

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:03:57 ; Search time 18 Seconds  
(without alignments)  
670.578 Million cell updates/sec

Title: US-09-466-778-11

Perfect score: 1857

Sequence: 1 MTGPGHKCKSHYVVDGL.....ALAAYSYFRINRKTIGFXHF 353

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36.\*

- 1: /SIDSL1/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDSL1/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDSL1/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDSL1/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDSL1/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDSL1/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDSL1/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDSL1/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDSL1/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDSL1/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDSL1/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDSL1/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDSL1/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDSL1/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDSL1/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDSL1/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDSL1/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDSL1/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDSL1/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDSL1/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDSL1/gcgdata/geneseq/geneseq/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262	14.1	277	13 R26044	Tumour necrosis fa
2	262	14.1	277	18 W13654	Tumour necrosis fa
3	262	14.1	277	20 W84087	Tumour necrosis fa
4	187.5	10.1	360	20 Y13381	Amino acid sequenc
5	169.5	9.1	339	21 Y57083	Rat proteoglycan 1
6	165.5	8.9	354	21 Y57080	Human link protein
7	165.5	8.9	355	21 Y57082	Chicken proteoglyc
8	163.5	8.8	354	21 Y57081	Human proteoglycan
9	163.5	8.8	354	21 Y57084	Human proteoglycan
10	157.5	8.5	1257	15 R46627	Neurocan core prot
11	156.5	8.4	252	19 W74523	Cloned duplicate o
12	155.5	8.4	355	16 R77034	Rat brain-enriched

13	155.5	8.4	912	16	R85442	Bovine brevicin co
14	155	8.3	528	19	W59882	Amino acid sequenc
15	151	8.1	116	19	W74522	Amino acid sequenc
16	150.5	8.1	2409	12	R12609	Version. Homo sa
17	148	8.0	908	16	R85443	Rat brevicin core
18	147	7.9	378	16	R77035	Cat brain-enriched
19	144	7.8	354	15	R57350	Murine delta prote
20	114.5	6.2	722	21	Y79028	Human cartilage li
21	109	5.9	322	20	Y13379	Amino acid sequenc
22	109	5.9	322	21	Y87287	Human signal pepti
23	107	5.8	322	19	W56249	Amino acid sequenc
24	107	5.8	2189	11	R05222	Antigen GX5401FL e
25	105	5.7	810	18	W37500	Human nel-related
26	99.5	5.4	727	18	W11719	C-Delta-1 polypept
27	99.5	5.4	740	18	W00876	C-Delta-1 polypept
28	98.5	5.3	728	21	Y79029	Chick delta protei
29	97.5	5.3	503	12	R14768	Metastasis-specifi
30	97	5.2	116	20	Y12323	Human 5' EST secre
31	96	5.2	361	13	R20816	Haematopoietic CD4
32	96	5.2	361	17	R91444	Human haematopoiet
33	96	5.2	361	19	W80453	Human CD44 antigen
34	96	5.2	361	20	W86200	Human CD44 antigen
35	96	5.2	493	13	R20817	Epithelial CD44 An
36	96	5.2	493	17	R91445	Human epithelial C
37	96	5.2	493	19	W80454	Human CD44 antigen
38	96	5.2	493	20	W89151	Human CD44 antigen
39	95.5	5.1	1480	13	R25079	Drosophila SLIT pr
40	94.5	5.1	722	18	W11720	M-Delta-1 polypept
41	93.5	5.0	512	20	W75494	Truncated human de
42	93.5	5.0	520	18	W18348	Proliferation and
43	93.5	5.0	702	18	W18349	Proliferation and
44	93.5	5.0	702	20	W75495	Truncated human de
45	93.5	5.0	723	18	W18353	Proliferation and

ALIGNMENTS

RESULT 1	
R26044	
ID R26044 standard; Protein; 277 AA.	
XX	
AC R26044;	
XX	
DT 03-FEB-1993 (first entry)	
XX	
DE Tumour necrosis factor-induced glycoprotein TSG-6.	
XX	
KW TNF; chronic inflammatory conditions; rheumatoid arthritis; sepsis;	
KW cancer; infections.	
XX	
OS Homo sapiens.	
XX	
FH Key	Location/Qualifiers
FT Peptide	1..19
FT	/*tag= a
FT	/note= "putative signal sequence"
FT	41..44
FT	/*tag= b
FT	/note= "potential N-glycosylation site"
FT	118..120
FT	/*tag= c
FT	/note= "potential N-glycosylation site"
FT	256..260
FT	/*tag= d
FT	/note= "potential N-glycosylation site"
XX	
PN W09212175-A.	
XX	
PD 23-JUL-1992.	
XX	
PF 14-JAN-1992;	92WO-US00333.
XX	





SQ Sequence 360 AA;

Query Match 10.1%; Score 187.5; DB 20; Length 360;  
 Best Local Similarity 30.8%; Pred. No. 1.8e-11;  
 Matches 45; Conservative 24; Mismatches 48; Indels 29; Gaps 5;

QY 5 GKHKCECKSHYVGDGLNCEPEQLPIDRCLQDNGOCHADAKCVDLHFQDITVGVFHLRSPL 64  
 Db 142 gryrce-----vidglede-----sglvelelrgv---vfpqspn 174  
 QY 65 GOYKLTDFDKAREACANEATMATYNQLSYQKAKYHLCAGWLETRGVAYPTAFASQNCG 124  
 Db 175 gryqfifhegqqvcaeqaavvasfelfaweegldwcnagwldqatvqyplmlprqpcg 234

QY 125 S-GVV-GIVDYGPRPNKSEMMDVFCY 148  
 Db 235 gpglapgvrsygprrhlrlhrydvfcf 260

RESULT 5  
 Y57083  
 ID Y57083 standard; protein; 339 AA.  
 AC Y57083;  
 DT 28-FEB-2000 (first entry)  
 XX XX  
 DE Rat proteoglycan link protein precursor.  
 XX XX  
 KW Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response;  
 KW immunotherapeutic agent; insulin dependent diabetes mellitus;  
 KW multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;  
 KW uveoretinitis; inflammatory response.  
 XX XX  
 OS Rattus sp.  
 XX XX  
 PN WO9956763-AL.  
 XX XX  
 PD 11-NOV-1999.  
 XX XX  
 PF 07-MAY-1999; 99WO-US10250.  
 XX XX  
 PR 07-MAY-1998; 98US-0084636.  
 XX XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX XX  
 PI Kaufman DL, Tian J, Olcott A;  
 XX XX  
 DR WPI; 2000-052905/04.  
 XX XX  
 PT Administration of neglected target tissue antigens to modulate immune  
 PT responses -  
 XX XX  
 PS Disclosure; Page 27; 79pp; English.  
 XX XX  
 CC Amino acid sequences Y57063-Y57091 are examples of neglected target  
 CC tissue antigens NNTAs. NNTAs are antigens (whole antigens or fragments)  
 CC not involved in autoimmunity. These peptides and proteins are used in the  
 CC method of the invention which involves administering an NNTA as an  
 CC antigen based immunotherapeutic agent, to a host afflicted with an  
 CC autoimmune response associated with an autoimmune disease. The  
 CC immunotherapeutic agent is used to treat autoimmune diseases such as  
 CC insulin dependent diabetes mellitus, multiple sclerosis, autoimmune  
 CC thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal inflammatory  
 CC immune responses. The NNTA induces regulatory tolerance by elicitation of  
 CC regulatory T cells among T cells recognizing the NNTA but not  
 CC participating in the immune response. The NNTA are capable of recognition  
 CC by substantial populations of uncommitted T cells which can be primed, or  
 CC biased, towards regulatory responses to provide effective treatment. The  
 CC NNTA are effective in regulating undesirable immune responses even when  
 CC target determinants used as agents promoting tolerance agents have failed  
 CC to induce an effective regulatory T cell response. NNTAs as agents

CC promoting tolerance are anticipated to be safer than use of target  
 CC determinants.

XX XX  
 SQ Sequence 339 AA;

Query Match 9.1%; Score 169.5; DB 21; Length 339;  
 Best Local Similarity 30.8%; Pred. No. 1.3e-09;  
 Matches 45; Conservative 20; Mismatches 52; Indels 29; Gaps 4;

QY 5 GKHKCECKSHYVGDGLNCEPEQLPIDRCLQDNGOCHADAKCVDLHFQDITVGVFHLRSPL 64  
 Db 120 grykce-----viegl-----edttavvalelqgv---vfpfprl 152  
 QY 65 GOYKLTDFDKAREACANEATMATYNQLSYQKAKYHLCAGWLETRGVAYPTAFASQNCG 124  
 Db 153 grylnfhearqacldqdviasfdqldawrgldwcnagwisdvsgvypitkprepg 212

QY 125 --SGVVGIVDYGPRPNKSEMMDVFCY 148  
 Db 213 gqntvpgrnygfwdkksrydvfcf 238

RESULT 6  
 Y57080  
 ID Y57080 standard; protein; 354 AA.  
 XX XX  
 AC Y57080;  
 DT 28-FEB-2000 (first entry)  
 XX XX  
 DE Human link protein precursor amino acid sequence.  
 XX XX  
 KW Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response;  
 KW immunotherapeutic agent; insulin dependent diabetes mellitus;  
 KW multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;  
 KW uveoretinitis; inflammatory response.  
 XX XX  
 OS Homo sapiens.  
 XX XX  
 PN WO9956763-AL.  
 XX XX  
 PD 11-NOV-1999.  
 XX XX  
 PF 07-MAY-1999; 99WO-US10250.  
 XX XX  
 PR 07-MAY-1998; 98US-0084636.  
 XX XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX XX  
 PI Kaufman DL, Tian J, Olcott A;  
 XX XX  
 DR WPI; 2000-052905/04.  
 XX XX  
 PT Administration of neglected target tissue antigens to modulate immune  
 PT responses -  
 XX XX  
 PS Disclosure; Page 26; 79pp; English.  
 XX XX  
 CC Amino acid sequences Y57063-Y57091 are examples of neglected target  
 CC tissue antigens NNTAs. NNTAs are antigens (whole antigens or fragments)  
 CC not involved in autoimmunity. These peptides and proteins are used in the  
 CC method of the invention which involves administering an NNTA as an  
 CC antigen based immunotherapeutic agent, to a host afflicted with an  
 CC autoimmune response associated with an autoimmune disease. The  
 CC immunotherapeutic agent is used to treat autoimmune diseases such as  
 CC insulin dependent diabetes mellitus, multiple sclerosis, autoimmune  
 CC thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal inflammatory  
 CC immune responses. The NNTA induces regulatory tolerance by elicitation of  
 CC regulatory T cells among T cells recognizing the NNTA but not  
 CC participating in the immune response. The NNTA are capable of recognition  
 CC by substantial populations of uncommitted T cells which can be primed, or  
 CC biased, towards regulatory responses to provide effective treatment. The  
 CC NNTA are effective in regulating undesirable immune responses even when  
 CC target determinants used as agents promoting tolerance agents have failed  
 CC to induce an effective regulatory T cell response. NNTAs as agents

CC NNTA are effective in regulating undesirable immune responses even when  
 CC target determinants used as agents promoting tolerance agents have failed  
 CC to induce an effective regulatory T cell response. NNTAs as agents  
 CC promoting tolerance are anticipated to be safer than use of target  
 CC determinants.

XX Sequence 354 AA;

Query Match 8.9%; Score 165.5; DB 21; Length 354;  
 Best Local Similarity 30.1%; Pred. No. 3.8e-09;  
 Matches 44; Conservative 22; Mismatches 51; Indels 29; Gaps 4;  
 QY 5 GKHKCECKSHYVGDGLNCEPEQLPDRCLQDNGOCHADAKCVDLHFQDTTGVVPHLSRPL 64  
 Db 135 grykce-----vieg|-----eddtavvaldlqgv---vfpypfprl 167  
 QY 65 GYKLTDFDKAREACANEATMATYNOLSYXOKAKYHLCISAGWLETGRVAYPTAFASQNCG 124  
 Db 168 grynlnfheaqaacldqdasfdqlydawrsigldwcnagwlsdgsqvypitkprepcg 227  
 QY 125 --SGVVGVIVDYGPRNPKSEMDVFCY 148  
 Db 228 gqntvpgvrnygfwdkdkrsydvfcf 253

RESULT 7

Y57082  
 ID Y57082 standard; protein; 355 AA.

XX AC Y57082;

XX DT 28-FEB-2000 (first entry)

XX DE Chicken proteoglycan link protein precursor.

XX KW Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response;  
 KW immunotherapeutic agent; insulin dependent diabetes mellitus;  
 KW multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;  
 KW uveoretinitis; inflammatory response.

XX OS Gallus sp.

XX PN WO9956763-Al.

XX PD 11-NOV-1999.

XX PF 07-MAY-1999; 99WO-US10250.

XX PR 07-MAY-1998; 98US-0084636.

XX PA (REGC ) UNIV CALIFORNIA.

XX PI Kaufman DL, Tian J, Olcott A;

XX XX WPI; 2000-052905/04.

XX DR Administration of neglected target tissue antigens to modulate immune  
 PT responses -

XX PS Disclosure; Page 27; 79pp; English.

XX CC Amino acid sequences Y57063-Y57091 are examples of neglected target  
 CC tissue antigens NNTAs. NNTAs are antigens (whole antigens or fragments)  
 CC not involved in autoimmunity. These peptides and proteins are used in the  
 CC method of the invention which involves administering an NNTA as an  
 CC antigen based immunotherapeutic agent, to a host afflicted with an  
 CC autoimmune response associated with an autoimmune disease. The  
 CC immunotherapeutic agent is used to treat autoimmune diseases such as  
 CC insulin dependent diabetes mellitus, multiple sclerosis, autoimmune  
 CC thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal inflammatory  
 CC immune responses. The NNTA induces regulatory tolerance by elicitation of  
 CC regulatory T cells among T cells recognizing the NNTA but not

CC participating in the immune response. The NNTA are capable of recognition  
 CC by substantial populations of uncommitted T cells which can be primed, or  
 CC biased, towards regulatory responses to provide effective treatment. The  
 CC NNTA are effective in regulating undesirable immune responses even when  
 CC target determinants used as agents promoting tolerance agents have failed  
 CC to induce an effective regulatory T cell response. NNTAs as agents  
 CC promoting tolerance are anticipated to be safer than use of target  
 CC determinants.

XX Sequence 355 AA;

Query Match 8.9%; Score 165.5; DB 21; Length 355;  
 Best Local Similarity 29.9%; Pred. No. 3.8e-09;  
 Matches 44; Conservative 26; Mismatches 46; Indels 31; Gaps 5;  
 QY 5 GKHKCECKSHYVGDGLNCEPEQLPDRCLQDNGOCHADAKCVDLHFQDTTGVVPHLSRSP- 63  
 Db 136 grykce-----vieg|-----eddtavvalnle---gvvfpyspr 167  
 QY 64 LGQYKLTDFDKAREACANEATMATYNOLSYXOKAKYHLCISAGWLETGRVAYPTAFASQNC 123  
 Db 168 lgrynlnfheaqaacldqdasfdqlyeawrsigldwcnagwlsdgsqvypitkprepc 227  
 QY 124 G--SGVVGVIVDYGPRNPKSEMDVFCY 148  
 Db 228 gqntvpgvrnygfwdkdkrsydvfcf 254

RESULT 8

Y57081

ID Y57081 standard; protein; 354 AA.

XX AC Y57081;

XX DT 28-FEB-2000 (first entry)

XX DE Human proteoglycan link protein precursor amino acid sequence.

XX KW Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response;  
 KW immunotherapeutic agent; insulin dependent diabetes mellitus;  
 KW multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;  
 KW uveoretinitis; inflammatory response.

XX OS Homo sapiens.

XX PN WO9956763-Al.

XX PD 11-NOV-1999.

XX PF 07-MAY-1999; 99WO-US10250.

XX PR 07-MAY-1998; 98US-0084636.

XX PA (REGC ) UNIV CALIFORNIA.

XX PI Kaufman DL, Tian J, Olcott A;

XX XX WPI; 2000-052905/04.

XX DR Administration of neglected target tissue antigens to modulate immune  
 PT responses -

XX PS Disclosure; Page 26-27; 79pp; English.

XX CC Amino acid sequences Y57063-Y57091 are examples of neglected target  
 CC tissue antigens NNTAs. NNTAs are antigens (whole antigens or fragments)  
 CC not involved in autoimmunity. These peptides and proteins are used in the  
 CC method of the invention which involves administering an NNTA as an  
 CC antigen based immunotherapeutic agent, to a host afflicted with an  
 CC autoimmune response associated with an autoimmune disease. The  
 CC immunotherapeutic agent is used to treat autoimmune diseases such as  
 CC insulin dependent diabetes mellitus, multiple sclerosis, autoimmune





XX WPI; 1998-449114/39.  
 DR N-PSDB; V53720; V53721.  
 XX

XX Production of homogeneous polysaccharides from heterogeneous  
 PT polysaccharides used for diagnosis and therapy of diseases -  
 PT comprises immobilisation on support e.g. polymer matrix and  
 PT selective cleavage with e.g. glucosidase or hydrolase  
 XX

XX Example 1; Fig 4; 19pp; English.

PS This is the amino acid sequence of the cloned duplicate of human  
 XX aggrecan G1-B domain, starting from the RBS of the tet o/p of  
 CC pr-HA-20, used in the method of the invention which involves the  
 CC production of homogeneous polysaccharides from heterogeneous  
 CC polysaccharides. The oligonucleotides and polysaccharides are  
 CC useful for the diagnosis and therapy of diseases.  
 XX

SQ Sequence 252 AA;

Query Match 8.4%; Score 156.5; DB 19; Length 252;  
 Best Local Similarity 25.9%; Pred. No. 2e-08;  
 Matches 49; Conservative 29; Mismatches 96; Indels 15; Gaps 4;

QY 57 VFHLRSPGQKLTDFDKAREACANEATMATYNQLSYXQKAKYHLCAGHLETRVAYPT 116  
 DB 31 VFHRAISRYTIDFDRACIQNSAIATPEQLQAYEDGFHQCDAGWLADQTVRYPI 90

QY 117 AFASQNC---GSGVGVGIVDYGRPNKSEMMDVFCY--RMKDVNCTKKVGVGDFGSYSGN 171  
 DB 91 HTPREGCYGDKDEFPGVRTYGIIR-DTNETYDVYCFAEELGTGRTGLEVVVKGI VFYRAI 149

QY 172 LLQVLMSPSLNPLTEVLAYSNSAGRAFLRH-----LTDLSIRGTLFXPONSNG 222  
 DB 150 STRYTLDFDRACQLNSAIATPEQLQAYEDGFHQCDAGWLADQTVRYPIHTPREG 209

QY 223 LGENETLSG 231

DB 210 YGDKDEFPG 218

RESULT 12  
 ID R77034 standard; Protein; 355 AA.  
 XX  
 AC R77034;

XX 28-APR-1996 (first entry)

XX Rat brain-enriched hyaluronan binding protein.

XX Brain-enriched hyaluronan binding protein; BEHAB; glioma;  
 KW brain tumour; hyaluronic acid; proteoglycan; tumour marker;  
 KW diagnostic.  
 XX

XX Rattus sp.

XX Key Location/Qualifiers

FT Peptide 1..22

FT Domain /label= Sig\_peptide

FT Domain 23..157

FT Domain /label= Ig\_fold\_domain

FT Domain 158..257

FT Domain /label= PTR1

FT Domain 258..355

FT Domain /label= PTR2

XX WO9527785-A1.

XX 19-OCT-1995.

XX 07-APR-1995; 95WO-US04353.

XX 08-APR-1994; 94US-0225477.  
 XX (UYVA ) UNIV YALE.  
 PA

XX Hockfield S, Jaworski DM;

XX WPI; 1995-366390/47.

DR N-PSDB; T01475.

XX Mammalian brain-enriched hyaluronan-binding protein and its DNA -  
 PT useful as diagnostic marker for detection of brain tumours and other  
 PT neuro: pathological states

XX Example 2; Fig 1; 46pp; English.

XX Novel rat brain-enriched hyaluronan binding (BEHAB) protein (R77034)  
 CC is isolated from brain tissue and has a high degree of homology to  
 CC members of the proteoglycan tandem repeat family of hyaluronan  
 CC binding proteins. It includes 2 domains, PTR1 and PTR2, thought  
 CC to be involved in hyaluronan binding. BEHAB is expressed at markedly  
 CC increased levels in glioma tissue, and can therefore be used as a  
 CC diagnostic marker. Recombinant BEHAB is obtd. by expression of the  
 CC isolated encoding cDNA sequence (T01475) in procaryotic or eucaryotic  
 CC host cells.  
 XX

SQ Sequence 355 AA;

Query Match 8.4%; Score 155.5; DB 16; Length 355;  
 Best Local Similarity 27.9%; Pred. No. 4.3e-08;  
 Matches 46; Conservative 21; Mismatches 61; Indels 37; Gaps 5;

QY 10 ECKSHVVDGLNCEPEQLPIDRCLQD-----NGOCHADAKCVLHF 50  
 DB 201 QCDAGWLSdqTVRYPIQNPACRYGMDGPGVGRNYGVGPDdlydvycyae----dln- 255

QY 51 QDQTVGVFHLRSLPGYKLTDFDKAREACANEATMATYNQLSYXQKAKYHLCAGHLETRG 110

DB 256 -----gelflgappg--kltweeardyclergaqiastgqlyaaawnggldrcspgwiadg 308

QY 111 RVAYPTAFASQNGSGVGVGIVDYGRPN-----KSEMMDVFCYR 149

DB 309 svrypiitpsqrcggcgpgvktlflfpnqtgfpksqnrnfvyvycfr 353

RESULT 13

ID R85442

XX R85442 standard; Protein; 912 AA.

XX AC R85442;

XX 19-FEB-1996 (first entry)

XX Bovine brevican core protein.

XX Brevican; chondroitin sulphate proteoglycan; glial cell; axon;  
 KW neurofibromatosis; gliosis.  
 XX

OS Bos taurus.

XX Key Location/Qualifiers

FT Peptide 1..34

FT Region /label= Sig\_peptide

FT Region 35..158

FT Modified-site /label= Immunoglobulin-like\_loop

FT Modified-site 130

FT Region /label= N-glycosylation\_site

FT Region 159..353

FT Modified-site /label= Protein-like\_tandem\_repeats

FT Modified-site 337

FT Region /label= N-glycosylation\_site

FT Region 354..648





```

PR 22-FEB-1997; 97EP-0102951.
XX
XX (LANS/) LANSING M.
PA (SCHM/) SCHMIDT G.
PA (UHLE/) UHLENKUEKEN J.
XX
XX LANSING M;
PI
XX
XX WPI; 1998-449114/39.
DR N-PSDB; V53715.
XX
XX Production of homogeneous polysaccharides from heterogeneous
PT polysaccharides used for diagnosis and therapy of diseases -
PT comprises immobilisation on support e.g. polymer matrix and
PT selective cleavage with e.g. glucosidase or hydrolase
XX
XX Example 1; Fig 3; 19pp; English.
PS
XX This is the amino acid sequence of the human aggrecan G1-B domain,
CC used in the method of the invention which involves the production
CC of homogeneous polysaccharides from heterogeneous polysaccharides.
CC The oligonucleotides and polysaccharides are useful for the diagnosis
CC and therapy of diseases.
XX
XX Sequence 116 AA;
SQ

Query Match 8.1%; Score 151; DB 19; Length 116;
Best Local Similarity 35.8%; Pred. No. 2.4e-08;
Matches 34; Conservative 15; Mismatches 42; Indels 4; Gaps 2;

QY 57 VFHLRSPLGQYKLTEDKAREACANEATMATYNQLSYXOKAKYHLCGAGWLETGRVAYPT 116
Db 14 VFHYRAISTRYTIDFDRACIQMSAIIATPEGLQAAYEDGFHQCDAGWIAQTVRYPI 73

QY 117 AFASQNC---GSGVYGIVDYGPRPNKSEMMDVFCY 148
Db 74 htpregcygdkdefpgvrtgyir-dtnetydvycf 107

```

Search completed: April 4, 2001, 13:05:01  
Job time: 64 sec

GenCore version 4.5  
 Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:04:01 ; Search time 27.49 Seconds  
 (without alignments)  
 1505.071 Million cell updates/sec

Title: US-09-466-778-11  
 Perfect score: 1857  
 Sequence: 1 MTGPGKHCKECSHYVGDGL.....ALAAYSYFRINKTIGFXHF 353

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SPTREMBL\_15:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_unclassified:\*
- 14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1757	94.6	897	Q9NRV3	Q9nr3 homo sapien
2	1468	79.1	1069	Q9UF98	Q9uf98 homo sapien
3	662	35.6	2212	Q33072	Q33072 homo sapien
4	645.5	34.8	2570	Q9NY15	Q9ny15 homo sapien
5	264	14.2	275	O08859	O08859 mus musculus
6	177	9.5	2109	P79787	P79787 gallus gall
7	169.5	9.1	355	O921X7	O921x7 mus musculus
8	162	8.7	1321	O44594	O44594 homo sapien
9	161.5	8.7	655	O08564	O08564 rattus norv
10	161.5	8.7	1290	Q9W6E1	Q9w6e1 gallus gall
11	160.5	8.6	2394	O77610	O77610 bos taurus
12	153	8.2	192	O02817	O02817 oryctolagus
13	153	8.2	656	O77612	O77612 bos taurus
14	151.5	8.2	1643	O77611	O77611 bos taurus
15	151.5	8.2	3381	O77609	O77609 bos taurus
16	126.5	6.8	103	Q9TBT3	Q9ttb3 sus scrofa
17	120.5	6.5	103	O46380	O46380 oryctolagus
18	110	5.9	396	Q9W6S4	Q9w6s4 gallus gall
19	109	5.9	302	Q9UNF4	Q9unf4 homo sapien

20	109	5.9	322	4	Q9F5Y7	Q9f5y7 homo sapien
21	105	5.7	1584	5	Q93791	Q93791 caenorhabdi
22	104	5.6	3507	5	Q23587	Q23587 caenorhabdi
23	103.5	5.6	717	13	P87357	P87357 brachydanio
24	102.5	5.5	778	13	Q9IBG4	Q9ibg4 xenopus lae
25	100	5.4	816	11	O70474	O70474 rattus norv
26	99.5	5.4	682	4	Q9NTM2	Q9ntm2 homo sapien
27	99.5	5.4	728	13	Q90556	Q90556 gallus gall
28	99.5	5.4	802	13	O57462	O57462 brachydanio
29	99.5	5.4	3623	4	O60494	O60494 homo sapien
30	99	5.3	780	11	Q9QX8	Q9qx8 mus musculu
31	99	5.3	1440	5	O20204	O20204 caenorhabdi
32	99	5.3	2506	11	Q9WUH9	Q9wuh9 rattus norv
33	98	5.3	294	4	Q92493	Q92493 homo sapien
34	98	5.3	2531	5	O16004	O16004 lytechinus
35	97.5	5.3	780	11	O08779	O08779 rattus norv
36	97.5	5.3	1480	5	Q9V7F8	Q9v7f8 drosophila
37	97.5	5.3	1504	5	Q9XYV4	Q9xyv4 drosophila
38	97.5	5.3	1504	5	Q9V7F9	Q9v7f9 drosophila
39	96	5.2	364	11	O70509	O70509 rattus norv
40	96	5.2	742	4	Q9UJ36	Q9uj36 homo sapien
41	95.5	5.1	364	6	O97569	O97569 ceratotheri
42	95.5	5.1	721	13	Q91902	Q91902 xenopus lae
43	95	5.1	261	10	O24530	O24530 vitis vinif
44	94.5	5.1	1328	13	P79754	P79754 fugu rubrip
45	94	5.1	637	10	Q9SF50	Q9sf50 arabidopsis

## ALIGNMENTS

RESULT 1

Q9NRV3 PRELIMINARY; PRT; 897 AA.

AC Q9NRV3; Q9NRV3; 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)

DE CD44-LIKE PRECURSOR FELL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Tao O., Zhang W., Cao X.;

RT "Molecular cloning and characterization of human FELL sharing homology with CD44.";

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF160476; AAF82398.1; -.

SQ SEQUENCE 897 AA; 97585 MW; EB920AF36101E388 CRC64;

Query Match 94.6%; Score 1757; DB 4; Length 897;  
 Best Local Similarity 93.8%; Pred. No. 8.6e-162;  
 Matches 331; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

QY 1 MTGPGKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL 60  
 |||||  
 DB 489 MTGPGKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL 548

QY 61 RSLPGQYKLTFFDKARACANEATMATYNQLSYOKAKYHLCSSAGWLETGRVAYPTAFAS 120  
 |||||  
 DB 549 RSLPGYKLTFFDKARACANEATMATYNQLSYOKAKYHLCSSAGWLETGRVAYPTAFAS 608

QY 121 QNCGSGVGVGVDPKPNKSEMDVFCYRMKDVNCTKXGVYVGDGFSYSGNLQVLMSEFP 180  
 |||||  
 DB 609 QNCGSGVGVGVDPKPNKSEMDVFCYRMKDVNCTKXGVYVGDGFSYSGNLQVLMSEFP 668

QY 181 SLTNFTEVLAYSNSARGRAFLHLEHTDLSIRGTLFXPONSGLGENETLSGRDIEHHLAN 240  
 |||||  
 DB 669 SLTNFTEVLAYSNSARGRAFLHLEHTDLSIRGTLFXPONSGLGENETLSGRDIEHHLAN 728

```
QY 241 VSMFFYNDLVNGTTLOPRGLSGKLLITDRQDPLHPTETRCVGDGRDTELDICASNGITHVI 300
Db 729 VSMFFYNDLVNGTTLOTRVSGSKLLITASQDPLQPTETRFVDGRAILQWDIFASNGIHHVI 788
QY 301 SRXLKAPPAPVTLXHTLXGIFXXIILVTGAVALAAYSFRINRRTIGFXXHF 353
Db 789 SRPLKAPPAPVTLTHTGLGAGIFFAILLVTGAVALAAYSFRINRRTIGFQHF 841

RESULT 2
Q9UF98 PRELIMINARY; PRT; 1069 AA.
AC OSUF98
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHETICAL 115.7 KDA PROTEIN (FRAGMENT).
GN DKF2P434B0321.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (K1AA0201-K1AA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
DR EMBL; D87433; BAA13377.1; -.
DR HSSP; P98066; ITSG.
DR INTERPRO; IPR000538; -.
DR EMBL; ALI33021; CAB61358.1; -.
DR HSSP; P98066; ITSG.
DR INTERPRO; IPR000538; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000782; -.
DR INTERPRO; IPR002049; -.
DR PFAM; PF00008; EGF_8.
DR PFAM; PF00193; Xlink; 1.
DR PRINTS; PR0193; Xlink; 1.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01265; LINKMODULE.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 5.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1069 AA; 115666 MW; 04B0960932164D63 CRC64;

Query Match 79.1%; Score 1468; DB 4; Length 1069;
Best Local Similarity 80.5%; Pred. No. 1.2e-133;
Matches 284; Conservative 4; Mismatches 29; Indels 36; Gaps 2;

QY 1 MTGPGKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDITVGVFHL 60
Db 697 MTGPGKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDITVGVFHL 756
QY 61 RSPGLQYKLTDFDKAREACANEATMATYNOLSYXQAKYHLCAGWLETGRVAYPTAFAS 120
Db 757 RSPGLQYKLTDFDKAREACANEATMATYNOLSYXQAKYHLCAGWLETGRVAYPTAFAS 816
QY 121 QNCGSGVGVIVDYGPRNPKSEMVDVFCYRMKDVNCTKKVGVGDGFSYSGNLLQVLMSP 180
Db 817 QNCGSGVGVIVDYGPRNPKSEMVDVFCYRMKG-----SAGLEFQQLSSRP 860
QY 181 SLTNFLTEVLAYSNSARGRAFLHLEHTDLSIRGLTFXPQNSGLGENETLSGRDIEHHLA 240
Db 861 CIS-----RTPDDLIRGLTFXPQNSGLGENETLSGRDIEHHLA 900
QY 241 VSMFFYNDLVNGTTLOPRGLSGKLLITDRQDPLHPTETRCVGDGRDTELDICASNGITHVI 300
Db 901 VSMFFYNDLVNGTTLOPRGLSGKLLITASQDPLQPTETRFVDGRAILQWDIFASNGIHHVI 960
QY 301 SRXLKAPPAPVTLXHTLXGIFXXIILVTGAVALAAYSFRINRRTIGFXXHF 353
Db 961 SRPLKAPPAPVTLTHTGLGAGIFFAILLVTGAVALAAYSFRINRRTIGFQHF 1013
```

```
RESULT 3
Q93072 PRELIMINARY; PRT; 2212 AA.
AC Q93072;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE MYELOBLAST KIAA0246 PROTEIN (FRAGMENT).
GN KIAA0246.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (K1AA0201-K1AA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
DR EMBL; D87433; BAA13377.1; -.
DR HSSP; P98066; ITSG.
DR INTERPRO; IPR000538; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000782; -.
DR INTERPRO; IPR001128; -.
DR INTERPRO; IPR002049; -.
DR PFAM; PF00008; EGF_13.
DR PFAM; PF00193; Xlink; 1.
DR PRINTS; PR01265; LINKMODULE.
DR PROSITE; PS00022; EGF_1; UNKNOWN_5.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 13.
DR PROSITE; PS01241; LINK; UNKNOWN_1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 2.
KW Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 2212 AA; 237451 MW; 4A95460504129134 CRC64;

Query Match 35.6%; Score 662; DB 4; Length 2212;
Best Local Similarity 40.2%; Pred. No. 4.9e-55;
Matches 143; Conservative 47; Mismatches 158; Indels 8; Gaps 5;

QY 2 TGPGRKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDITVGVFHL 60
Db 1796 TGLNTRRCECHAGYVGDGLQCLESEPPVDRCLGQPPPCHSAMDCTDLHFQEKRAGVFHL 1855
QY 61 RSPGLQYKLTDFDKAREACANEATMATYNOLSYXQAKYHLCAGWLETGRVAYPTAFAS 120
Db 1856 QATSPGYGLNFSEAEACAEAGAVLASPPQLSAAQQLGFLCLMGWLANGSTAHPVVPV 1915
QY 121 QNCGSGVGVIVDYGPRNPKSEMVDVFCYRMKDVNCTKKVGVGDGFSYSGNLLQVLMSP 179
Db 1916 ADCGNRGVIVSLGARKNLSERWDAYCFRQDVACRCRNGFVGDICTCNKGLDLVAAT 1975
QY 180 PSLTNFLTEVLAYSNSARGRAFLHLEHTDLSIRGLTFXPQNSGLGENETLSGRDIEHHLA 239
Db 1976 ANFSTFYGMLLGYANATQRLDLDLDELTYKTLFVPVNEGFVDMNLTSGSPDLEHAS 2035
QY 240 NVSMFFYNDLVNGTTLOPRGLSGKLLITDR-----QDPLHPTETRCVGDGRDTELDICASNGI 296
Db 2036 NATLISAN-ASQKLLPAHSGLSLISDAGPDNNSWAPVCTVVVSRIIIVMDIMAFNGI 2094
QY 297 THVISRXLKAPPAPVTLXHTLXGIFXXIILVTGAV--ALAAYSYFRINRRTIGF 350
Db 2095 IHALASPLAPPQPAVLAPEAPPVAAVGAVLAAGALLGLVAGALYLRRACKPMGF 2150
```

RESULT	4		
ID	Q9NY15	PRELIMINARY;	PRT; 2570 AA.
AC	Q9NY15;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)		
DE	STABLIN-1.		
OS	STAB1.		
GN	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Pollitz O., Guillot P., Gratchev A., Schledzewski K., Birk R.,		
RA	Hakiv N., Tebbe B., Orfanos C.E., Goerd S.,		
RT	"Stabilin-1: an endothelial-macrophage member of the fasciclin domain		
RT	containing protein family associated with angiotensins."		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ275213; CAB61827.1; -.		
SQ	SEQUENCE 2570 AA: 275346 MW; 3123FABD7C8E2BF8 CRC64;		
<p>Query Match 34.8%; Score 645.5; DB 4; Length 2570;            Best Local Similarity 39.3%; Pred. No. 2.4e-53;            Matches 142; Conservative 48; Mismatches 154; Indels 17; Gaps</p>			
QY	2	TGPGKHCKECKSHYVGDLNC-EPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL	60
DB	2153	TGLNTRRCECHAGYVGDLQCLUEESEPVDRLGIGLQPPCHSDAMCTDQHFQEKRAGVFHL	2212
QY	61	RSLPGQVKLTFRKARACANEATMATYNQLSYOKAKYHLCSAGWLETGRVAVYPTAFAS	120
DB	2213	QATSGPGYGLNFSAEACAEQAGVIAFSLQLSAAQLGFHLGMLGWLANGSTAHVPVFPV	2272
QY	121	QNCGSGYGVIVDYGPRPNKSEMDVFCYRMKDVNCTXKYYVGDGFS-YSGNLLQVLMSE	179
DB	2273	ADCGNGRVGVVSLGARKNLSEWDAYCFRVQDVACRCNGFVGDSIGTSCNGKLLDLAAT	2332
QY	180	PSLTNFIETVLAYSNSARGRAFLHLTDLSTIRGTLFXPQNSGLGENTTLSGRDIEHHLA	239
DB	2333	ANFSTFGMLLYANATQRLGDLFLDDELTYKTLFPVNEGFDVNMNLTSGPDLHLHAS	2392
QY	240	NVSMFEFVNDLVNCTTLQTRLGSKLITDR--QDPLHPTETRCVDGRDTLSEWDICASNGI	296
DB	2393	NATLLSAN-ASQKLLPAHSGLSLIISADGPNSSWAPVAGTGVVVSRIIVIMAFNGI	2451
QY	297	THVISRLKAPPAPVTLX-----HTXLGXGIFXXIILVTGAV--ALAAYSYFRINRKTIG	349
DB	2452	IHALASPELLAPPQAVLAPEAPPAAGVG----AVLAAGALLGLVAGALLRARGRTG	2507
QY	350	F 350	
DB	2508	F 2508	
RESULT	5		
ID	O08859	PRELIMINARY;	PRT; 275 AA.
AC	O08859;		
DT	01-JUL-1997 (TrEMBLrel. 04, Created)		
DT	01-JUL-1997 (TrEMBLrel. 04, Last sequence update)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)		
DE	TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN TSG-6 (HYALURONATE-BINDING		
DE	PROTEIN).		
GN	TNFI6 OR TNFAIP6 OR TSG6.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		

RX	MEDLINE=98087423; PubMed=9427551;
RA	Fulop C., Kamath R.V., Li Y., Otto J.M., Salustri A., Olsen B.R.,
RB	Giant T.T., Hascall V.C.;
RT	"Coding sequence, exon-intron structure and chromosomal localization
RT	of murine TNF-stimulated gene 6 that is specifically expressed by
RT	expanding cumulus cell-oocyte complexes.";
RL	Gene 202:95-102(1997).
DR	EMBL; U83903; AAC53527.1; -.
DR	HSSP; P98066; ITSG.
DR	MGD; MGI:1195266; Tnfip6.
DR	INTERPRO; IPR000538; -.
DR	INTERPRO; IPR000859; -.
DR	PFAM; PF00193; Xlink; 1.
DR	PFAM; PF00431; CUB; 1.
DR	PROSITE; PS01180; CUB; 1.
DR	PROSITE; PS01241; LINK; 1.
DR	PRODOM; PD000918; -, 1.
SQ	SEQUENCE 275 AA; 30924 MW; 1CD247228260B8F9 CRC64;
Query Match 14.2%; Score 264; DB 11; Length 275;	
Best Local Similarity 43.3%; Pred.No.1.2e-17;	
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps	
Qy	52 DTTGVFHLRSLGQYKLTFKAREACANEATWATYNQLSYQKAKYHLCISAGWLETGR 11
Dd	: 11:: : 11::: :: : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd	32 EQAAGVYHREARAGRYKLTYAEAKAVCFEFGGLRATVKQLAAARKIGFHVCAGHWAKGR 91
Qy	112 VAYPTAFASONGCGVGIVDYGPPNKSEMMVDVFCYRMKDVMC 155
Dd	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd	92 VGYPVIRKPGNGCFGKTGIIDYGINLRSERWDACYNPHAKEC 135
RESULT 6	
P79787	
ID	P79787 PRELIMINARY; PRT: 2109 AA.
AC	C79787;
DT	01-MAY-1997 (TREMBLrel. 03, Created)
DT	01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE	CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN.
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae
OC	Gallus.
OX	NCBI_TaxID=9031;
ON	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=BRAIN;
RX	MEDLINE=96262324; PubMed=8965652;
RA	Li H., Domowicz M., Hennig A., Schwartz N.B.;
RT	"S103L reactive chondroitin sulfate proteoglycan (aggrecan) mRNA
RT	expressed in developing chick brain and cartilage is encoded by a
RT	single gene.";
RL	Brain Res. Mol. Brain Res. 36:309-321(1996).
DR	EMBL; U78555; AAC60053.1; -.
DR	HSSP; P08709; IBF9.
DR	INTERPRO; IPR000152; -.
DR	INTERPRO; IPR000436; -.
DR	INTERPRO; IPR000538; -.
DR	INTERPRO; IPR000561; -.
DR	INTERPRO; IPR000742; -.
DR	INTERPRO; IPR001304; -.
DR	INTERPRO; IPR001881; -.
DR	INTERPRO; IPR003006; -.
DR	PFAM; PF00008; EGF; 1.
DR	PFAM; PF00047; ig; 1.
DR	PFAM; PF00059; lectin_c; 1.
DR	PFAM; PF00084; sushi; 1.
DR	PFAM; PF00193; Xlink; 4.
DR	PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR	PROSITE; PS00615; C TYPE LECTIN 1; 1.



```

QY 55 VGVFHLRSLGOYKLTDFDKAREACANEATMATYNQLSYXOKAKYHLCSAGWLETGRVAY 114
Db 159 GVVFHRSARDRYALTFAEQAACRLSIAIAAPRHQAFAEDGFDNCDAGWLSDRTVRY 218
QY 115 PTFASQNC---GSGVVGTVDYGPRNKSEMDVFCYRKMKNVCNTXKVGYYG 163
Db 219 PITQSRPGCYGRSSLPGRVSYG-RRNPQELIDVYCFARE---LGGEVYVG 266

RESULT 9
ID O88564 PRELIMINARY; PRT; 655 AA.
AC O88564;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE VERISCAN V3 ISOFORM PRECURSOR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=WISTAR KYOTO;
RA Lemire J.M., Braun K.R., Maurel P., Margolis R.U., Schwartz S.M.,
RA Wight T.N.;
RT "Vertican isoforms in vascular smooth muscle cells.";
RL EMBL: AF072892; AAC26116.1; -;
DR HSSP; P01132; 1EGF.
DR INTERPRO: IPR000152; -;
DR INTERPRO: IPR000436; -;
DR INTERPRO: IPR000538; -;
DR INTERPRO: IPR000561; -;
DR INTERPRO: IPR000742; -;
DR INTERPRO: IPR001304; -;
DR INTERPRO: IPR001438; -;
DR INTERPRO: IPR001881; -;
DR INTERPRO: IPR003006; -;
DR PFAM; PF00008; EGF_2; -;
DR PFAM; PF00047; Ig; 1; -;
DR PFAM; PF00059; lectin_c; 1;
DR PFAM; PF00084; sushi; 1;
DR PFAM; PF00193; Xlink; 2;
DR PRINTS; PR00010; EGFBL00D.
DR PROSITE; PS00010; ASX HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
DR PRODOM; PD000918; -; 2.
KW Signal; Glycoprotein; EGF-like domain.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 655 VERSICAN V3 ISOFORM.
SQ SEQUENCE 655 AA; 74474 MW; 602F2F37E4F1BCC4 CRC64;

Query Match 8.7%; Score 161.5; DB 11; Length 655;
Best Local Similarity 26.1%; Pred. No. 3.5e-07;
Matches 73; Conservative 30; Mismatches 98; Indels 79; Gaps 13;

QY 57 VFHLRSLGOYKLTDFDKAREACANEATMATYNQLSYXOKAKYHLCSAGWLETGRVAYPT 116
Db 151 VFHYRAATSYRTLNFEQAQACIDIGAVIATPELFAAAYEDGFEQCDAGWLSQDTVRYP 210
QY 117 AFASQNCGS---GVGVGVYDYGPRNKSEMDVFCYRKMKNVCNTXKVGYYG-GDGFYSGNL 172
Db 211 RAPREGCYGDMGKGVRYGFR-SQETIYDVICY-----VDHLDGVFVH----- 254
QY 173 LQVLMSPSLTNFLTEVLAYSNSARGRAFLE-----HLTDLSIRG--TL 215

```

```

Db 255 ----ITAPSKFTFEAEAEACANRDLARLTVGELHAARNGFDQCQDYGWLSDASVRRHPVT 310
QY 216 FXPNQSG--IG-----ENETL----SGRDIERHLANYSMFFYNDLVNGTTLTQTRLGSK 262
Db 311 ARACCGGGLLVRTLYRFENQTCFPLPDSRFDAYCFKRPDLCKTNPCLNGGT----- 362
QY 263 LLITRDQDPLHPTET-----RCVDG----RDTLEWDICASN 294
Db 363 -----CYPTETSYVCTCAPGYSGDQCELDLDFDECHSN 393

RESULT 10
Q9W6E1 PRELIMINARY; PRT; 1290 AA.
ID Q9W6E1;
AC Q9W6E1;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE NEUROCAN CORE PROTEIN PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RA Li H., Leung T., Balsamo J., Hoffman S., Lillien J.;
RT "cDNA cloning of chicken neurocan and its role in regulating N-
cadherin function in embryonic chicken retina.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF116856; AAD24546.2; -;
DR HSSP; P08709; 1BF9.
DR INTERPRO: IPR000152; -;
DR INTERPRO: IPR000436; -;
DR INTERPRO: IPR000538; -;
DR INTERPRO: IPR000561; -;
DR INTERPRO: IPR000742; -;
DR INTERPRO: IPR001304; -;
DR INTERPRO: IPR001438; -;
DR INTERPRO: IPR001881; -;
DR PFAM; PF00008; EGF_2; -;
DR PFAM; PF00059; lectin_c; 1;
DR PFAM; PF00084; sushi; 1;
DR PFAM; PF00193; Xlink; 2;
DR PRINTS; PR00010; EGFBL00D.
DR PROSITE; PS00010; ASX HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 1290 AA; 138876 MW; 182BD86D0E40BE78 CRC64;

Query Match 8.7%; Score 161.5; DB 13; Length 1290;
Best Local Similarity 29.1%; Pred. No. 8.6e-07;
Matches 43; Conservative 21; Mismatches 51; Indels 33; Gaps 5;

QY 5 GKHKCECKSHYVGLNCEPEQLIDRCLODNGOCHADAKCYDLHFQDTTVG-VFHLRSP 63
Db 136 GLYRCE-----VVAGIDDDSDLLPLE-----VMGVVFHYRPA 167
QY 64 LGQYKLTDFDKAREACANEATMATYNQLSYXOKAKYHLCSAGWLETGRVAYPTAFASNC 123
Db 168 GARYALTFAAARRACMNSAVIASPOHLQAPEFDGVDNCDAGWLDQSVRYPTILSRPGC 227
QY 124 ---GSGVGVYDYGPRNKSEMDVFCY 148
Db 228 YGDRNSLPGVRSYQGR-EPGELYDVICY 254

```

```
RESULT 11
O77610 PRELIMINARY; PRT; 2394 AA.
AC O77610;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE VERSICAN V1 SPLICE-VARIANT PRECURSOR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9828320; Pubmed=9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain."
RL J. Biol. Chem. 273:15758-15764 (1998).
DR EMBL; AF060457; AAC24359.1; -
DR HSSP; P01132; LEFG.
DR INTERPRO; IPR000152; -
DR INTERPRO; IPR000436; -
DR INTERPRO; IPR000538; -
DR INTERPRO; IPR000561; -
DR INTERPRO; IPR000742; -
DR INTERPRO; IPR000742; -
DR INTERPRO; IPR001304; -
DR INTERPRO; IPR001438; -
DR INTERPRO; IPR001881; -
DR INTERPRO; IPR003006; -
DR PFAM; PF00008; EGF; 2.
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sushi; 1.
DR PFAM; PF00084; sushi; 1.
DR PFAM; PF00193; Xlink; 2.
DR PRINTS; PRO0010; EGFLOOD.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
DR PRODOM; PD000918; -; 2.
KW Signal; Glycoprotein; EGF-like domain.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 2394 VERSICAN V1 SPLICE-VARIANT.
SQ SEQUENCE 2394 AA; 261886 MW; B82A3E10FC5BD990 CRC64;

Query Match 8.6%; Score 160.5; DB 6; Length 2394;
Best Local Similarity 25.5%; Pred. No. 2.4e-06;
Matches 83; Conservative 31; Mismatches 113; Indels 99; Gaps 17;

Qy 51 QDT---TVG---VFLHSLPGQYKLTEDKAREACANEATATYNQLSYXQAKYHLCSAG 105
Db 141 QDTVSLTVEGVFHYRAATSRVTLNFEAMQKACVDICAVIATPEQLHAAAYEDGECQDAG 200
Qy 106 WLETGRVAYPTAFASQNGS---GVGVGIVDYGPRPNKSEMDVFCY-----148
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 WLSQTVRYPIRVPRGECYDMMKGEGVRYTGFPRA-PHETVDYICVVDHLDGVDVHTAP 259
Qy 149 -----RMKDVNCTXKGVYGD-----GF-----S 167
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 260 NKFTFEAGEECKTQD----ARLATVGEQLQAAWRNGFDRCDYGWLLDASVRHPTVVARAQ 315
Qy 168 YSNNLQV--LMFSPSLTNFLTEVLAYSNSARGAFLEHLTDLSTRGTLFXPQNSGLGE 225
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 CGGGLLGVRTLYRFENQTGFPT-----PDSRFDAYCFKRMSDFSVSG---HPIDSESKE 367

Query Match 8.2%; Score 153; DB 6; Length 192;
Best Local Similarity 35.8%; Pred. No. 4.5e-07;
Matches 34; Conservative 15; Mismatches 42; Indels 4; Gaps 2;

Qy 57 VFHLRSLPGQYKLTEDKAREACANEATATYNQLSYXQAKYHLCSAGHLETGRVAYPT 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 VFHYRAISTRYTLDFDRAQRAQLNSAIATPEQLQAAAYEDGFGHCDAGWLADQTVRYPI 156
Qy 117 AFASQNC---GSVVGVGIVDYGPRPNKSEMDVFCY 148
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 157 HTPREGCGYKDEFFGVRTYIGIR-DTNETYDVICYF 190

RESULT 12
O02817 PRELIMINARY; PRT; 192 AA.
ID O02817;
AC O02817;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE LARGE AGGREGATING CARTILAGE PROTEOGLYCAN CORE PROTEIN (FRAGMENT).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CARTILAGE;
RA Nishimura M., Noshiro M., Kawamoto T., Nakamasu K., Hamada T.,
RA Kato Y.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB004812; BAA20524.1; -
DR HSSP; P98066; 1TSG.
DR INTERPRO; IPR000538; -
DR INTERPRO; IPR003006; -
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00193; Xlink; 1.
DR PRINTS; PR01265; LINKMODULE.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE; PS01241; LINK; 1.
FT NON_TER 1 192
FT NON_TER 192 192
SQ SEQUENCE 192 AA; 21606 MW; 44EA35FA92CEB8CC CRC64;

Query Match 8.2%; Score 153; DB 6; Length 192;
Best Local Similarity 35.8%; Pred. No. 4.5e-07;
Matches 34; Conservative 15; Mismatches 42; Indels 4; Gaps 2;

Qy 57 VFHLRSLPGQYKLTEDKAREACANEATATYNQLSYXQAKYHLCSAGHLETGRVAYPT 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 VFHYRAISTRYTLDFDRAQRAQLNSAIATPEQLQAAAYEDGFGHCDAGWLADQTVRYPI 156
Qy 117 AFASQNC---GSVVGVGIVDYGPRPNKSEMDVFCY 148
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 157 HTPREGCGYKDEFFGVRTYIGIR-DTNETYDVICYF 190

RESULT 13
O77612 PRELIMINARY; PRT; 656 AA.
ID O77612;
AC O77612;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE VERSICAN V3 SPLICE-VARIANT PRECURSOR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9828320; Pubmed=9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain."
RL J. Biol. Chem. 273:15758-15764 (1998).
DR EMBL; AF060457; AAC24359.1; -
DR HSSP; P01132; LEFG.
DR INTERPRO; IPR000152; -
DR INTERPRO; IPR000436; -
DR INTERPRO; IPR000538; -
DR INTERPRO; IPR000561; -
DR INTERPRO; IPR000742; -
DR INTERPRO; IPR000742; -
DR INTERPRO; IPR001304; -
DR INTERPRO; IPR001438; -
DR INTERPRO; IPR001881; -
DR INTERPRO; IPR003006; -
DR PFAM; PF00008; EGF; 2.
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sushi; 1.
DR PFAM; PF00084; sushi; 1.
DR PFAM; PF00193; Xlink; 2.
DR PRINTS; PRO0010; EGFLOOD.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
DR PRODOM; PD000918; -; 2.
KW Signal; Glycoprotein; EGF-like domain.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 2394 VERSICAN V1 SPLICE-VARIANT.
SQ SEQUENCE 2394 AA; 261886 MW; B82A3E10FC5BD990 CRC64;
```



DR EMBL; AF060456; AAC24358.1; -.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:04:21 ; Search time 11.97 Seconds  
(without alignments)  
952.365 Million cell updates/sec

Title: US-09-466-778-11  
Perfect score: 1857  
Sequence: 1 MTGPGKHCKECKSHYVGDGL.....ALAAYSYFRINKTIGFXHF 353

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	264	14.2	276	1	TSG6_RABIT	P98065 oryctolagus
2	262	14.1	277	1	TSG6_HUMAN	P98066 homo sapien
3	175	9.4	2109	1	PGCA_CHICK	P07898 gallus gall
4	171	9.2	2124	1	PGCA_RAT	P07897 rattus norv
5	170	9.2	2364	1	PGCA_BOVIN	P13608 bos taurus
6	169.5	9.1	354	1	PLK_RAT	P03994 rattus norv
7	169.5	9.1	356	1	PLK_MOUSE	Q9qup5 mus musculu
8	166.5	9.0	354	1	PLK_HORSE	Q28381 equus cabal
9	165.5	8.9	354	1	PLK_BOVIN	P52522 bos taurus
10	165.5	8.9	355	1	PLK_CHICK	P07354 gallus gall
11	165	8.9	2132	1	PGCA_MOUSE	Q61282 mus musculu
12	163.5	8.8	354	1	PLK_HUMAN	P10915 homo sapien
13	161.5	8.7	354	1	PLK_PIG	P10859 sus scrofa
14	161.5	8.7	2415	1	PGCA_HUMAN	P16112 homo sapien
15	158.5	8.5	1268	1	PGCN_MOUSE	P55066 mus musculu
16	157.5	8.5	883	1	PGCB_RAT	P55068 rattus norv
17	157.5	8.5	1257	1	PGCN_RAT	P55067 rattus norv
18	155.5	8.4	394	1	PGCA_RABIT	Q28670 oryctolagus
19	155.5	8.4	912	1	PGCB_BOVIN	Q28062 bos taurus
20	154.5	8.3	862	1	PGCV_MACNE	Q28858 macaca neme
21	152.5	8.2	3358	1	PGCV_MOUSE	Q62059 mus musculu
22	152.5	8.2	3562	1	PGCV_CHICK	Q20953 gallus gall
23	150.5	8.1	883	1	PGCB_MOUSE	Q61361 mus musculu
24	150.5	8.1	3396	1	PGCV_HUMAN	P13611 homo sapien
25	147	7.9	417	1	PGCB_FELCA	P41725 felis silve
26	111.5	6.0	362	1	CD44_CRIGR	P20944 cricetus
27	108.5	5.8	431	1	CD44_MESAU	Q60522 m cd44 anti
28	106	5.7	359	1	CD44_HORSE	Q05078 equus cabal
29	105	5.7	810	1	NELL_HUMAN	Q28232 homo sapien
30	103.5	5.6	655	1	CD44_MOUSE	P15379 mus musculu
31	100	5.4	351	1	CD44_CANFA	Q28284 canis fami
32	99	5.3	2907	1	FBN2_MOUSE	Q61555 mus musculu
33	98	5.3	742	1	CD44_HUMAN	P16070 h cd44 anti

34	97.5	5.3	503	1	CD44_RAT	P26051 rattus norv
35	97.5	5.3	1480	1	SLIT_DROME	P24014 drosophila
36	97	5.2	2911	1	FBN2_HUMAN	P35556 homo sapien
37	96	5.2	713	1	TSA4_GIALA	P21849 giardia lam
38	95.5	5.1	366	1	CD44_BOVIN	Q29423 bos taurus
39	95.5	5.1	683	1	BGH3_PIG	O11780 sus scrofa
40	94.5	5.1	722	1	DL11_MOUSE	Q61483 mus musculu
41	94.5	5.1	1375	1	NID2_HUMAN	Q14112 homo sapien
42	93.5	5.0	723	1	DL11_HUMAN	Q00548 homo sapien
43	93	5.0	1245	1	NIDO_MOUSE	P10493 mus musculu
44	92.5	5.0	362	1	CD44_PAPHA	P14745 papio hamad
45	92.5	5.0	1403	1	NID2_MOUSE	O88322 mus musculu

## ALIGNMENTS

RESULT 1  
TSG6\_RABIT  
ID TSG6\_RABIT STANDARD; PRT; 276 AA.  
AC P98065;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN TSG-6 PRECURSOR (HYALURONATE-  
DE BINDING PROTEIN PS4)  
GN TNFAIP6 OR TSG6 OR PS4.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Lagomorpha; Leporidae; Oryctolagus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NEW ZEALAND WHITE;  
RX MEDLINE=93252803; PubMed=8098034;  
RA Feng P., Liau G.;  
RT "Identification of a novel serum and growth factor-inducible gene in  
RT vascular smooth muscle cells.";  
RL J. Biol. Chem. 268:9387-9392(1993).  
RN [2]  
RP ERATUM.  
RX MEDLINE=94012707; PubMed=8407990;  
RA Feng P., Liau G.;  
RL J. Biol. Chem. 269:21453-21453(1993).  
CC -1- FUNCTION: POSSIBLY INVOLVED IN CELL-CELL AND CELL-MATRIX  
CC INTERACTIONS DURING INFLAMMATION AND TUMORGENESIS.  
CC -1- TISSUE SPECIFICITY: VASCULAR SMOOTH MUSCLE CELLS.  
CC -1- DEVELOPMENTAL STAGE: FETAL.  
CC -1- INDUCTION: BY SERUM AND GROWTH FACTOR.  
CC -1- SIMILARITY: CONTAINS 1 LINK DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M86381; AAA03342.1; -  
CC HSSP; P98066; 1TSG.  
DR INTERPRO; IPR000538; -  
DR INTERPRO; IPR000859; -  
DR PFAM; PF00431; CUB; 1.  
DR PFAM; PF00193; Xlink; 1.  
DR PROSITE; PS01180; CUB; 1.  
DR PROSITE; PS01241; LINK; 1.  
KW Cell adhesion; Signal; Glycoprotein.  
FT SIGNAL 1 19  
FT CHAIN 20 276  
FT DOMAIN 53 128  
FT LINK.  
FT DOMAIN 135 247

```
FT DISULFID 58 127 BY SIMILARITY.
FT DISULFID 82 103 BY SIMILARITY.
FT DISULFID 135 161 BY SIMILARITY.
FT DISULFID 188 210 BY SIMILARITY.
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 276 AA; 31081 MW; 3BDC5D9A24B2F75A CRC64;

Query Match 14.2%; Score 264; DB 1; Length 276;
Best Local Similarity 43.3%; Pred. No. 2.5e-16;
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

Qy 52 DTTVGVEHLRSPGLQYKLTDFKAREACANEATWATYNQLSYXQKAKYHLCAGWLETGR 111
Db 32 EQAAGVYHREARSGKYKLTAEAKAVCEFEGLRLATYKQLEAAARKIGFHVCAAGWMAKGR 91

Qy 112 VAYPTAFASONGSGVGVGDYGRPNKSEMDVFCYRMKDVC 155
Db 92 VGYPIVAPGSCGFGKTIIDYGIRLNRSEMDAYCYNPHAKEC 135

RESULT 2
TSG6_HUMAN STANDARD; PRT; 277 AA.
AC P98066;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN TSG-6 PRECURSOR (HYALURONATE-
DE BINDING PROTEIN).
GN TNFAIP6 OR TSG6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP "A novel secretory tumor necrosis factor-inducible protein (TSG-6) is
RP a member of the family of hyaluronate binding proteins, closely
RP related to the adhesion receptor CD44."
RL J. Cell Biol. 116:545-557(1992).
RN [2]
RP STRUCTURE BY NMR OF 36-133.
RX MEDLINE=96390850; PubMed=8797823;
RA Kohda D., Morton C.J., Parker A.A., Hatanaka H., Inagaki F.M.,
RA Campbell I.D., Day A.J.;
RA "Solution structure of the link module: a hyaluronan-binding domain
RT involved in extracellular matrix stability and cell migration."
RL Cell 86:767-775(1996).
CC -!- FUNCTION: POSSIBLY INVOLVED IN CELL-CELL AND CELL-MATRIX
CC INTERACTIONS DURING INFLAMMATION AND TUMORGENESIS.
CC -!- INDUCTION: BY TNF.
CC -!- SIMILARITY: CONTAINS 1 LINK DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31165; AAB00792.1; -.
DR PDB; 1TSG; 01-APR-97.
DR MM; 600410.
DR INTERPRO; IPR000538; -.
DR INTERPRO; IPR000859; -.
DR PFM; PF00431; CUB; 1.
DR PFM; PF00193; Xlink; 1.
```

```
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01241; LTNK; 1.
KW Cell adhesion; Signal; Glycoprotein; 3D-structure.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 277 TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN
FT TSG-6.
FT DOMAIN 53 128 LINK.
FT DOMAIN 135 247 CUB.
FT DISULFID 58 127
FT DISULFID 82 103
FT DISULFID 135 161 BY SIMILARITY.
FT DISULFID 188 210 BY SIMILARITY.
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 277 AA; 31231 MW; 4DB3AEB4AC52B880 CRC64;

Query Match 14.1%; Score 262; DB 1; Length 277;
Best Local Similarity 43.3%; Pred. No. 3.7e-16;
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

Qy 52 DTTVGVEHLRSPGLQYKLTDFKAREACANEATWATYNQLSYXQKAKYHLCAGWLETGR 111
Db 32 ERAAGVYHREARSGKYKLTAEAKAVCEFEGLRLATYKQLEAAARKIGFHVCAAGWMAKGR 91

Qy 112 VAYPTAFASONGSGVGVGDYGRPNKSEMDVFCYRMKDVC 155
Db 92 VGYPIVAPGSCGFGKTIIDYGIRLNRSEMDAYCYNPHAKEC 135

RESULT 3
PGCA_CHICK STANDARD; PRT; 2109 AA.
ID PGCA_CHICK
AC P07898; Q90991; Q90820; Q91047; Q90810;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
DE PROTEIN) (CSPCP).
GN AGCI.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEHORN; TISSUE=EMBRYO;
RX MEDLINE=94043149; PubMed=8226878;
RA Li H., Schwartz N.B., Vertel B.M.;
RT "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core
RT protein and identification of a stop codon in the aggrecan gene
RT associated with the chondrodystrophy, nanomelia."
RL J. Biol. Chem. 268:23504-23511(1993).
RN [2]
RP SEQUENCE OF 1042-1559 FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE=90307744; PubMed=1694853;
RA Krueger R.C. Jr., Fields T.A., Mensch J.R. Jr., Schwartz N.B.;
RT "Chick cartilage chondroitin sulfate proteoglycan core protein. II.
RT Nucleotide sequence of cDNA clone and localization of the S103L
RT epitope."
RL J. Biol. Chem. 265:12088-12097(1990).
RN [3]
RP SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.
RC TISSUE=CARTILAGE;
RX MEDLINE=93111968; PubMed=1339285;
RA Chandrasekaran L., Tanzer M.L.;
RT "Molecular cloning of chicken aggrecan. Structural analyses."
RL Biochem. J. 288:903-910(1992).
RN [4]
RP ERRATUM.
RX MEDLINE=94107258; PubMed=8280087;
RA Chandrasekaran L., Tanzer M.L.;
```

RL	Biochem. J. 296:885-887(1993).	DR	INTERPRO: IPR001881; -.
RN	[5]	DR	PFAM: PF00008; EGF; 1.
RP	SEQUENCE OF 1492-1610 FROM N.A.	DR	PFAM: PF00193; Xlink; 4.
RC	STRAIN-WHITE LEHORN; TISSUE=CHONDROCYTES;	DR	PFAM: PF00059; lectin_c; 1.
RX	MEDLINE=95128519; PubMed=7827752;	DR	PFAM: PF00084; sushi; 1.
RA	Primorac D., Stover M.L., Clark S.H., Rowe D.W.;	DR	PROSITE: PS00010; ASX_HYDROXYL; 1.
RT	"Molecular basis of nanomelia, a heritable chondrodystrophy of	DR	PROSITE: PS00022; EGF_1; 1.
RT	chicken.";	DR	PROSITE: PS01187; EGF_CA; 1.
RT	Matrix Biol. 14:297-305(1994).	DR	PROSITE: PS01241; LINK; 4.
RL	[6]	DR	PROSITE: PS00615; C-TYPE LECTIN_1; 1.
RN	SEQUENCE OF 1894-2109 FROM N.A.	DR	PROSITE: PS50041; C-TYPE LECTIN_2; 1.
RP	MEDLINE=89008500; PubMed=3170613;	KW	Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
RA	Tanaka T., Har-El R., Tanzer M.L.;	KW	EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.
RT	"Partial structure of the gene for chicken cartilage proteoglycan	FT	POTENTIAL.
RT	core protein.";	FT	SIGNAL 1 16
RT	J. Biol. Chem. 263:15831-15835(1988).	FT	CHAIN 17 2109
RL	[7]	FT	DOMAIN 44 136
RN	SEQUENCE OF 1693-1855 AND 1893-2109 FROM N.A.	FT	DOMAIN 166 243
RP	MEDLINE=86259736; PubMed=3460082;	FT	DOMAIN 264 346
RA	Sai S., Tanaka T., Koshier R.A., Tanzer M.L.;	FT	DOMAIN 537 614
RT	"Cloning and sequence analysis of a partial cDNA for chicken	FT	DOMAIN 635 716
RT	cartilage proteoglycan core protein.";	FT	REPEAT 1363 1742
RT	Proc. Natl. Acad. Sci. U.S.A. 83:5081-5085(1986).	FT	DOMAIN 1855 1892
CC	-!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR	FT	DOMAIN 1901 2019
CC	MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN	FT	REPEAT 2023 2081
CC	IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO	FT	DOMAIN 48 137
CC	HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A	FT	DOMAIN 148 243
CC	REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.	FT	DOMAIN 249 346
CC	-!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY	FT	DOMAIN 519 613
CC	SIMILARITY).	FT	DOMAIN 620 715
CC	-!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO	FT	DOMAIN 718 803
CC	TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,	FT	DOMAIN 805 1264
CC	MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS	FT	DOMAIN 1265 1742
CC	CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS	FT	DOMAIN 1893 2109
CC	THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)	FT	DOMAIN 51 129
CC	AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2	FT	DISULFID 171 242
CC	AND G3.	FT	DISULFID 195 216
CC	-!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE	FT	DISULFID 269 345
CC	CHAINS. N-LINKED AND O-LINKED OLIGOSACCHARIDES.	FT	DISULFID 293 314
CC	-!- DISEASE: DEFECTS IN AGC1 IS THE CAUSE OF NANOMELIA, A LETHAL	FT	DISULFID 542 613
CC	CONNECTIVE TISSUE DISORDER AFFECTING CARTILAGE DEVELOPMENT	FT	DISULFID 566 587
CC	(CHONDRODYSTROPHY) CHARACTERIZED BY SHORTENED AND MALFORMED LIMBS.	FT	DISULFID 640 715
CC	AGGREGAN IS TRUNCATED AT IS C-TERMINAL IN THE CS-2 BINDING DOMAIN	FT	DISULFID 664 685
CC	AND IS NOT ANYMORE SECRETED FROM THE CHONDROCYTES.	FT	DISULFID 1859 1870
CC	-!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.	FT	DISULFID 1864 1879
CC	-!- SIMILARITY: CONTAINS 4 LINK DOMAINS.	FT	DISULFID 1881 1890
CC	-!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.	FT	DISULFID 1897 1908
CC	-!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.	FT	DISULFID 1925 2017
CC	-!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.	FT	DISULFID 1993 2009
CC	-!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.	FT	DISULFID 2024 2067
CC	-----	FT	DISULFID 2053 2080
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	FT	CARBOHYD 76 76
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	FT	CARBOHYD 122 122
CC	the European Bioinformatics Institute. There are no restrictions on its	FT	CARBOHYD 330 330
CC	use by non-profit institutions as long as its content is in no way	FT	CARBOHYD 388 388
CC	modified and this statement is not removed. Usage by and for commercial	FT	CARBOHYD 439 439
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	FT	CARBOHYD 644 644
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	FT	CARBOHYD 700 700
CC	-----	FT	CARBOHYD 765 765
CC	EMBL; L21913; AAB19128.1; -.	FT	CARBOHYD 801 801
DR	EMBL; M38187; AAA48731.1; -.	FT	VARSPPLIC 1856 1892
DR	EMBL; M88101; -; NOT_ANNOTATED_CDS.	FT	CONFLICT 362 362
DR	EMBL; S74657; AAC60751.1; -.	FT	CONFLICT 601 601
DR	EMBL; S74656; AAC60751.1; JOINED.	FT	CONFLICT 1000 1000
DR	EMBL; J04028; AAA48719.1; -.	FT	CONFLICT 1029 1029
DR	EMBL; M13993; AAA48720.1; -.	FT	CONFLICT 1042 1043
DR	PIR; A25442; A25442.	FT	CONFLICT 1251 1251
DR	HSSP; P00740; 1IXA.	FT	CONFLICT 1587 1587
DR	INTERPRO: IPR000152; -.	FT	CONFLICT 1590 1590
DR	INTERPRO: IPR000436; -.	FT	CONFLICT 1594 1594
DR	INTERPRO: IPR000538; -.	FT	CONFLICT 1602 1610
DR	INTERPRO: IPR000561; -.	FT	CONFLICT

FT CONFLICT 1988 1988 F -> S (IN REF. 6).  
SQ SEQUENCE 2109 AA; 223492 MW; 7F824FD5B3A2ABDA CRC64;

Query Match 9.4%; Score 175; DB 1; Length 2109;  
Best Local Similarity 23.4%; Pred. No. 2.4e-07;  
Matches 75; Conservative 31; Mismatches 116; Indels 98; Gaps 12;

QY 10 ECKSHYVDGLNCEPEQLPDRCLQDNGQCHA-----DAKVDLHFQDPTVG 56  
DB 194 QCDAGLWLDQVRRPIILPRRCVGRGDEFFGVRTYGVRETDFDYVICYAEQMGK--- 250  
QY 57 VFHLRSLPGQKLFDFKAREACAEATMATYNLSYXQAKYHLCAGWLETGRVAYPT 116  
DB 251 VFYATSP---EKFTFQAFDKCHSLGARLATTGELYLAWKDMDCSAGWLADBSVRYPI 307  
QY 117 AFASONGCGSVVGI--VDYGRP-----PNKSEMDVFCYRKMKNVC-----YXKVG-YV 162  
DB 308 SRARPNCGGNLGVRTVYLPANQTVGPHPSRYDAICYSGDDPEALVPGLFTEDEVCTEL 367  
QY 163 GDGFSY-----SCNLLQVLMSPPSLTNFL 186  
DB 368 GSAFTIQTQVTEVELPLPRNVNTEEEARGSIATLEPMEITATATLEYAEFTVLPDL--FA 425  
QY 187 TEVLAYSNSARGRAFLHLTDL-----SIRGTLFXPQNSGLGE----- 225  
DB 426 TSVIVETASPREENVNTRTEETGCIWAVPEVTVTSVSGTAF---TTGMAEVSSVEEAIAYTA 482  
QY 226 ---NETLSGRDIEHHLANVS 242  
DB 483 TPGLESASPTIEDHLVQVT 502

RESULT 4  
PGCA\_RAT STANDARD; PRT; 2124 AA.  
AC P07897;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE  
DE PROTEIN) (CSFPCP).  
GN AGC1 OR AGC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8087070; PubMed=3693370;  
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;  
RT "Complete primary structure of the rat cartilage proteoglycan core  
RT protein deduced from cDNA clones.";  
RL J. Biol. Chem. 262:17757-17767(1987).  
RN [2]  
RP REVISION TO 698.  
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;  
RL J. Biol. Chem. 263:10040-10040(1988).  
RN [3]  
RP SEQUENCE OF 1856-2124 FROM N.A.  
RX MEDLINE=86250698; PubMed=2424893;  
RA Doege K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;  
RT "Partial cDNA sequence encoding a globular domain at the C terminus  
RT of the rat cartilage proteoglycan.";  
RL J. Biol. Chem. 261:8108-8111(1986).  
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR  
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN  
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO  
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A  
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.  
CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY  
CC • SIMILARITY).  
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO  
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,  
CC

CC CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS  
CC CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS  
CC CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)  
CC CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2  
CC CC AND G3.  
CC CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE  
CC CC CHAINS. N-LINKED AND O-LINKED OLIGOSACCHARIDES.  
CC CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.  
CC CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.  
CC CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC CC the European Bioinformatics Institute. There are no restrictions on its  
CC CC use by non-profit institutions as long as its content is in no way  
CC CC modified and this statement is not removed. Usage by and for commercial  
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC CC or send an email to license@isb-sib.ch).  
CC CC -----  
CC CC EMBL: M13518; AAA41836.1; -;  
CC CC EMBL: J03485; AAA21000.1; ALT\_SEQ.  
CC CC PIR: A23835; A23835.  
CC CC PIR: A28452; A28452.  
CC CC HSSP: P98066; LTSG.  
CC CC INTERPRO: IPR000436; -;  
CC CC INTERPRO: IPR000495; -;  
CC CC INTERPRO: IPR000538; -;  
CC CC INTERPRO: IPR001304; -;  
CC CC PFAM: PF00193; Xlink; 4.  
CC CC PFAM: PF00059; Lctin.c; 1.  
CC CC PFAM: PF00084; sush1; 1.  
CC CC PROSITE: PS00290; IG\_MHC; 1.  
CC CC PROSITE: PS01241; LINK; 4.  
CC CC PROSITE: PS00615; C-TYPE\_LECTIN\_1; 1.  
CC CC PROSITE: PS00411; C-TYPE\_LECTIN\_2; 1.  
CC CC Repeat; Immunoglobulin domain.  
CC CC Repeat; Immunoglobulin domain.  
CC CC SIGNAL 1 19 POTENTIAL.  
CC CC CHAIN 20 2124 AGGREGAN CORE PROTEIN.  
CC CC DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.  
CC CC DOMAIN 170 247 LINK 1.  
CC CC DOMAIN 268 349 LINK 2.  
CC CC DOMAIN 504 581 LINK 3.  
CC CC DOMAIN 602 683 LINK 4.  
CC CC DOMAIN 1910 2036 C-TYPE LECTIN.  
CC CC DOMAIN 2040 2098 SUSHI.  
CC CC DOMAIN 48 140 GI-A.  
CC CC DOMAIN 152 247 GI-B.  
CC CC DOMAIN 253 349 GI-B'.  
CC CC DOMAIN 486 580 G2-B.  
CC CC DOMAIN 587 682 G2-B'.  
CC CC DOMAIN 685 798 KS.  
CC CC DOMAIN 801 1226 CS-1.  
CC CC DOMAIN 1227 1909 CS-2.  
CC CC DOMAIN 1910 2124 G3.  
CC CC DISULFID 51 133 BY SIMILARITY.  
CC CC DISULFID 175 246 BY SIMILARITY.  
CC CC DISULFID 199 220 BY SIMILARITY.  
CC CC DISULFID 273 348 BY SIMILARITY.  
CC CC DISULFID 297 318 BY SIMILARITY.  
CC CC DISULFID 509 580 BY SIMILARITY.  
CC CC DISULFID 533 554 BY SIMILARITY.  
CC CC DISULFID 607 682 BY SIMILARITY.  
CC CC DISULFID 631 652 BY SIMILARITY.  
CC CC DISULFID 1914 1925 BY SIMILARITY.  
CC CC DISULFID 1942 2034 BY SIMILARITY.  
CC CC DISULFID 2010 2026 BY SIMILARITY.  
CC CC DISULFID 2041 2084 BY SIMILARITY.  
CC CC DISULFID 2070 2097 BY SIMILARITY.  
CC CC CARBOHYD 126 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).









```
Db 135 GRYKCE-----VIEGL-----EDDTAVVALDLQGV---VFPYFRL 167
QY 65 GOYKLTEDKAREACANEATMATYNOLSYXQKAKYHLCSAGWLETGRVAYPTAFASQNGC 124
Db 168 GRYNLNFHEAQAQCLDQDAVIASFQDLYDAWRGLDWCNAGWLSGDSGVQYPTTKPREPCG 227
QY 125 --SGVVGVGDYDGRPNKSEMDVFCY 148
Db 228 GQNTVPGVRNYGFWDKESRYDVFCF 253

RESULT 9
PLK_BOVIN
ID PLK_BOVIN STANDARD; PRT; 354 AA.
AC P55252;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).
GN CPTL1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CARTILAGE;
RX MEDLINE=96043243; PubMed=7584851;
RA Hering T.M., Kollar J., Huynh T.D., Sandell L.J.;
RT "Bovine chondrocyte link protein cDNA sequence: interspecies
RT conservation of primary structure and mRNA untranslated regions.";
RL Comp. Biochem. Physiol. 112B:197-203(1995).
CC -!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH
CC HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U02292; AAC04311.1; -
CC HSP; P98066; ITSG.
CC INTERPRO; IPR000538; -
CC INTERPRO; IPR003006; -
CC PFAM; PF00193; Xlink; 2.
CC PFAM; PF00047; ig; 1.
CC PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;
KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.
FT PROPEP 1 15 BY SIMILARITY.
FT CHAIN 16 354 PROTEOGLYCAN LINK PROTEIN.
FT DOMAIN 54 146 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 176 253 LINK 1.
FT DOMAIN 274 350 LINK 2.
FT DISULFID 61 139 BY SIMILARITY.
FT DISULFID 181 252 BY SIMILARITY.
FT DISULFID 205 226 BY SIMILARITY.
FT DISULFID 279 349 BY SIMILARITY.
FT DISULFID 304 325 BY SIMILARITY.
SQ SEQUENCE 354 AA; 40287 MW; 065D155378A1283C CRC64;

Query Match 8.9%; Score 165.5; DB 1; Length 354;
Best Local Similarity 30.1%; Pred. No. 1.9e-07;
Matches 44; Conservative 22; Mismatches 51; Indels 29; Gaps 4;

QY 5 GKHKCEKSHYVGDLNCEPEQLPIDRLQDNGQCHADAKCVDLHFQDTTVGVFHLRSLP 64
```

```
Db 135 GRYKCE-----VIEGL-----EDDTAVVALDLQGV---VFPYFRL 167
QY 65 GOYKLTEDKAREACANEATMATYNOLSYXQKAKYHLCSAGWLETGRVAYPTAFASQNGC 124
Db 168 GRYNLNFHEAQAQCLDQDAVIASFQDLYDAWRGLDWCNAGWLSGDSGVQYPTTKPREPCG 227
QY 125 --SGVVGVGDYDGRPNKSEMDVFCY 148
Db 228 GQNTVPGVRNYGFWDKESRYDVFCF 253

RESULT 10
PLK_CHICK
ID PLK_CHICK STANDARD; PRT; 355 AA.
AC P07354;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).
GN CRTLL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYONIC STERNAL CARTILAGE;
RX MEDLINE=86233315; PubMed=3459154;
RA Deak F., Kiss I., Sparks K.J., Argaves W.S., Hampikian G.,
RA Goetnick P.F.;
RT "Complete amino acid sequence of chicken cartilage link protein
RT deduced from cDNA clones.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3766-3770(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87317659; PubMed=3476955;
RA Kiss I., Deak F., Mestric S., Dellius H., Soos J., Dekany K.,
RA Argaves W.S., Sparks K.J., Goetnick P.F.;
RT "Structure of the chicken link protein gene: exons correlate with the
RT protein domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6399-6403(1987).
CC -!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH
CC HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M35038; AAA48941.1; -
CC EMBL; M35035; AAA48941.1; JOINED.
CC EMBL; M35036; AAA48941.1; JOINED.
CC EMBL; M35037; AAA48941.1; JOINED.
CC EMBL; M13212; AAA48940.1; -
CC PIR; A24881; LKCH.
CC PIR; A28305; A28305.
CC HSP; P98066; ITSG.
CC INTERPRO; IPR000538; -
CC INTERPRO; IPR003006; -
CC PFAM; PF00193; Xlink; 2.
CC PFAM; PF00047; ig; 1.
CC PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;
KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.
FT PROPEP 1 9
FT CHAIN 10 355 PROTEOGLYCAN LINK PROTEIN.
FT DOMAIN 54 147 IG-LIKE V-TYPE DOMAIN.
```



This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

-----  
EMBL; X17405; CAA35462.1; -;  
EMBL; U43328; AAA85216.1; -;  
PIR; S14914; LKHU.  
PIR; A36308; A36308.  
HSSP; P98056; LTSG.  
MIM; I15435; -;  
INTERPRO; IPR000538; -;  
INTERPRO; IPR003006; -;  
PFAM; PF00193; Xlink; 2.  
PFAM; PF00047; ig; 1.  
PROSITE; PS01241; LINK; 2.  
KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;  
KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.  
FT PROPEP 1 15  
FT CHAIN 16 354  
FT DOMAIN 54 146  
FT DOMAIN 176 253  
FT DOMAIN 274 350  
FT DISULFID 61 139  
FT DISULFID 181 252  
FT DISULFID 205 226  
FT DISULFID 279 349  
FT DISULFID 304 325  
FT CARBOHYD 21 21  
FT CARBOHYD 56 56  
SEQUENCE 354 AA; 40165 MW; 315C96EC3AC2626A CRC64;  
  
Query Match 8.8%; Score 163.5; DB 1; Length 354;  
Best Local Similarity 30.1%; Pred. No. 2.8e-07;  
Matches 44; Conservative 21; Mismatches 52; Indels 29; Gaps 4;  
  
QY 5 GKHKCECKSHYYGDGINCEPEQLPIDRLCDLNQGOCHADAKVDLHFQDTVGVFHLRSP 64  
|::||| |::|| |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||  
Db 135 GRKYE-----VIEG-----EDTVVVALDIQGV---VFYPFR 167  
  
QY 65 GOYKLTDFKARCAEANEATMATYNQLSYXQAKYHLCISAGWLETRAVPTAFASQCNG 124  
|::||| |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||  
Db 168 GRYNLFHEAQACLDDAVIASPDQLYDAWRGLDCNAGWSDGSQVPITKPREPG 227  
  
QY 125 --SGVVIGVIDYGPKNKSEMMDVFCY 148  
: |: :|| :||||:  
Db 228 QONTVEGVGRNYGFWDKRSYDVFCF 253  
  
RESULT 13  
PLK\_PIG  
ID PLK\_PIG STANDARD; PRT; 354 AA.  
AC P10859;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).  
GN CRTLL.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
[1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LARYNGEAL CARTILAGE CHONDROCYTES;  
RX MEDLINE=89293837; PubMed=2738916;  
RA Perkins S.J., Nealis A.S., Dudha J., Hardingham T.E.;  
RT "Immunoglobulin fold and tandem repeat structures in proteoglycan N-terminal domains and link protein".  
RL J. Mol. Biol. 206:737-753(1989).  
RN



DR PFAM: PF000059; lectin\_c; 1.  
DR PFAM: PF000084; sushi; 1.  
DR PROSITE: PS000022; EGF\_1; 1.  
DR PROSITE: PS01186; EGF\_2; 1.  
DR PROSITE: PS00290; IG\_MHC; 1.  
DR PROSITE: PS01241; LINK; 3.  
DR PROSITE: PS00615; C\_TYPE\_LLECTIN\_1; 1.  
DR PROSITE: PS00411; C\_TYPE\_LLECTIN\_2; 1.  
KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;  
KW EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.  
FT SIGNAL 1 19  
FT CHAIN 20 2415  
FT DOMAIN 44 140  
FT DOMAIN 170 247  
FT DOMAIN 268 349  
FT DOMAIN 495 572  
FT DOMAIN 593 673  
FT DOMAIN 2164 2199  
FT DOMAIN 2201 2327  
FT REPEAT 2331 2389  
FT DOMAIN 48 141  
FT DOMAIN 152 247  
FT DOMAIN 253 349  
FT DOMAIN 477 571  
FT DOMAIN 578 672  
FT DOMAIN 676 848  
FT DOMAIN 772 844  
FT DOMAIN 851 1497  
FT DOMAIN 941 1497  
FT DOMAIN 1498 2162  
FT DOMAIN 2163 2415  
FT DISULFID 51 133  
FT DISULFID 175 246  
FT DISULFID 199 220  
FT DISULFID 273 348  
FT DISULFID 297 318  
FT DISULFID 500 571  
FT DISULFID 524 545  
FT DISULFID 598 672  
FT DISULFID 621 642  
FT DISULFID 2168 2178  
FT DISULFID 2173 2187  
FT DISULFID 2189 2198  
FT DISULFID 2205 2216  
FT DISULFID 2233 2325  
FT DISULFID 2301 2317  
FT DISULFID 2332 2337  
FT DISULFID 2361 2388  
FT CARBOHYD 126 126  
FT CARBOHYD 239 239  
FT CARBOHYD 333 333  
FT CARBOHYD 387 387  
FT CARBOHYD 434 434  
FT CARBOHYD 602 602  
FT CARBOHYD 657 657  
FT CARBOHYD 737 737  
FT CARBOHYD 1898 1898  
FT VARSPIC 2163 2200  
FT VARSPIC 2330 2390  
FT CONFLICT 766 766  
FT CONFLICT 847 847  
FT CONFLICT 1928 1928  
FT CONFLICT 1964 1964  
FT CONFLICT 2070 2070  
FT CONFLICT 2391 2391  
SQ SEQUENCE 2415 AA; 1288937E1B98C5B6 CRC64;  
  
Query Match 8.7%; Score 161.5; DB 1; Length 2415;  
Best Local Similarity 27.9%; Pred. No. 4.6e-06;  
Matches 50; Conservative 22; Mismatches 64; Indels 43; Gaps 6;  
  
47 DLHFQDRTVG-----VFHLRSLPGQYKLTDFDKAREACANEATMATYNLSYXQA 97

Db 460 DLVVQVTAVFQGPPLPGVVFHYRPGTPRYSLSLTFEEAQAACPGTGAIVASPEQLQAAYEA 519  
Qy 98 KYHLCSAGWLETGRVAVPTAFASQNC---GSGVVGIVDYGPRPNKSEMMWDFCY----- 148  
Db 520 GYEOCDAGLRDQVRPIVSPRTPCVGDKDSSPGVTVGRFS-TETVDYVCFVDRLEG 578  
Qy 149 -----RWKDV-----NCTXKVGIVGDFGS-----YSGNLQVLSMFFSLT 183  
Db 579 EVFFATRLQEQTFQEALEFCFESHNATATTQLYAANWSRGLDKCYAGWLADSLRYPITV 637  
  
RESULT 15  
PGCN\_MOUSE  
ID PGCN\_MOUSE STANDARD; PRT; 1268 AA.  
AC P55066;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE NEUROCAN CORE PROTEIN PRECURSOR.  
GN CSPG3 OR NCAN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALE/C; TISSUE=BRAIN;  
RX MEDLINE=96039250; PubMed=7490074;  
RA Rauch U., Grimpe B., Kulbe G., Arnold-Ammer I., Beier D.,  
RA Faessler R.;  
RL "Structure and chromosomal localization of the mouse neurocan gene."; Genomics 28:405-410(1995).  
CC -!- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NG-CAM AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC ACID.  
CC -!- TISSUE SPECIFICITY: BRAIN.  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.  
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X84727; CAA59216.1; -  
DR HSSP; P00740; LIXA.  
DR MGD; MGI:104694; CSPG3.  
DR INTERPRO; IPR000152; -  
DR INTERPRO; IPR000436; -  
DR INTERPRO; IPR000538; -  
DR INTERPRO; IPR000561; -  
DR INTERPRO; IPR001304; -  
DR INTERPRO; IPR001881; -  
DR INTERPRO; IPR003006; -  
DR PFAM; PF00008; EGF; 2.  
DR PFAM; PF00047; Ig; 1.  
DR PFAM; PF00059; lectin\_c; 1.  
DR PFAM; PF00084; sushi; 1.  
DR PROSITE; PS00010; ASK\_HYDROXYL; 1.  
DR PROSITE; PS00022; EGF\_1; 3.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 1.  
DR PROSITE; PS01241; LINK; 2.  
DR PROSITE; PS00615; C\_TYPE\_LLECTIN\_1; 1.

Search completed: April 4, 2001, 13:06:41  
Job time: 140 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:03:57 ; Search time 17.49 Seconds

(without alignments)  
1370.436 Million cell updates/sec

Title: US-09-466-778-11

Perfect score: 1857  
Sequence: 1 MTGPGKHCKESHVGDGL.....ALAAYSYFRINRKTIGFXHF 353

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_66:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1468	79.1	1069	T42681	hypothetical prote
2	264	14.2	275	JC6506	tumor necrosis fac
3	264	14.2	276	A47290	TSG-6 homolog P54
4	262	14.1	277	A41735	hyaluronate-bindin
5	175	9.4	2109	I50421	aggreacan precursor
6	171	9.2	2124	A28452	proteoglycan core
7	170	9.2	2327	T42630	aggreacan - bovine
8	169.5	9.1	408	LKRT2	proteoglycan link
9	166.5	9.0	354	S42938	proteoglycan link
10	165.5	8.9	355	LKCH	proteoglycan link
11	165	8.9	2132	A55182	aggreacan precursor
12	163.5	8.8	354	LKHU	proteoglycan link
13	161.5	8.7	354	S04243	proteoglycan link
14	161.5	8.7	2316	A39086	aggreacan precursor
15	160	8.6	370	S29139	aggreacan - pig (fr
16	158.5	8.5	1268	S52781	neurocan - mouse
17	157.5	8.5	371	A53908	brevican precursor
18	157.5	8.5	883	S49126	brevican precursor
19	157.5	8.5	1257	S28764	neurocan precursor
20	155.5	8.4	912	A54423	brevican precursor
21	154.5	8.3	862	S43922	versican - pig-tai
22	152.5	8.2	3562	A47171	chondroitin sulfat
23	151.5	8.2	1643	T14274	versican precursor
24	151.5	8.2	3381	T42389	versican precursor
25	150.5	8.1	883	S57653	brevican precursor
26	150.5	8.1	2409	A60979	versican precursor
27	148	8.0	2397	A55535	versican precursor
28	147	7.9	378	T146268	brevican precursor
29	139.5	7.5	1340	A39808	proteoglycan core

30	122	6.6	113	2	A55885	chondroitin sulfat
31	111.5	6.0	362	2	A35616	T-cell surface gly
32	109.5	5.9	358	2	A34907	plasma membrane gl
33	109.5	5.9	363	2	A37009	CD44 homolog membr
34	108	5.8	395	2	I77371	CD44R5 - human
35	106	5.7	359	2	S24240	lymphocyte surface
36	105	5.7	1584	2	T22674	hypothetical prote
37	104	5.6	3507	2	T34513	hypothetical prote
38	103.5	5.6	365	2	A34424	CD44 membrane gly
39	100	5.4	351	2	S45305	CD44 antigen precu
40	99.5	5.4	728	2	I50719	C-Delta-1 - chicke
41	99.5	5.4	3623	2	T09456	intrinsic factor-B
42	99	5.3	601	2	T22025	hypothetical prote
43	99	5.3	2907	2	A57278	fibrillin-2 precu
44	98	5.3	294	2	A32377	lymphocyte surface
45	98	5.3	361	2	G02251	cell surface glyco

ALIGNMENTS

RESULT 1

T42681  
hypothetical protein DKF2p434E0321.1 - human (fragments)  
C:Species: Homo sapiens (man)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000  
C:Accession: T42681  
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, November 1999  
A:Reference number: Z22233  
A:Accession: T42681  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-870:871-1069 <AAA>  
A:Cross-references: EMBL:AL133021  
A:Experimental source: adult testis; clone DKF2p434E0321  
A:Note: the cDNA sequence contains a +1 frameshift near codon 870  
C:Genetics:  
A:Note: DKF2p434E0321.1

Query Match	79.1%	Score	1468	DB	2	Length	1069
Best Local Similarity	80.5%	Pred. No.	1.1e-117				
Matches	284	Conservative	4	Mismatches	29	Indels	36
Gaps	2						
Qy	1	MTGPGKHCKESHVGDGLNCEPEQLPDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL	60				
Db	697	MTGPGKHCKESHVGDGLNCEPEQLPDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL	756				
Qy	61	RSPGLGQYKLTDFDKAREACANEATMATYNQLSYQKAKYHLCAGWLETGRVAYPTAFAS	120				
Db	757	RSPGLGQYKLTDFDKAREACANEATMATYNQLSYQKAKYHLCAGWLETGRVAYPTAFAS	816				
Qy	121	QNCSSGVVGVLDXGPRNKSEMDVFCYRMKDVNCTKVGYYGDSYSGNLQVLMSEF	180				
Db	817	QNCSSGVVGVLDXGPRNKSEMDVFCYRMKDVNCTKVGYYGDSYSGNLQVLMSEF	860				
Qy	181	SLTNFLTEVLAYSNSSARGRAFLHLDLSIRGTLFXPQNSGLGENETLSGRDIEHHLAN	240				
Db	861	CIS-----RTPDDLIRGTLFVPQNSGLGENETLSGRDIEHHLAN	900				
Qy	241	VSMEFFNDLVNGTTLQRLGSKLLITDRDPLHPTETRCVDCGRDITLWDICASNGITHVI	300				
Db	901	VSMEFFNDLVNGTTLQRLGSKLLITASODPLOTETRFVDGRRAILQWDIFASNGIILHVI	960				
Qy	301	SRXLKAPPAPVTLXHTXKLGXGIFXXIILVTGAVALAAYSIFRINRKTIGFXHF	353				
Db	961	SRPLKAPPAPVTLTHTGLGAGIFFAILVLTGAVALAAYSIFRINRRTIGFQHF	1013				

RESULT 2

JC6506  
tumor necrosis factor stimulated gene-6 protein - mouse

C:Species: Mus musculus (house mouse)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: J06506  
R:Fueloep, C.; Kamath, R.V.; Li, Y.; Otto, J.M.; Salustri, A.; Olsen, B.R.; Glant, T.T.;  
Gene 202, 95-102, 1997  
A:Title: Coding sequence, exon-intron structure and chromosomal localization of murine  
A:Reference number: J06506; MUID:98087423  
A:Accession: J06506  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-275 <FUE>  
A:Cross-references: GB:U83903; NID:g2062474; PIDN:AA03527.1; PID:g2062475

Query Match 14.2% Score 264; DB 2; Length 275;  
Best Local Similarity 43.3% Pred. No. 3.1e-15;  
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

Qy 52 DTTGVFHLRSPGLQYKLTDFDKAREACANEATMATYNQLSYXOKAKYHLCSSAGWLETGR 111  
Db 32 EQAAGVYHREARSGYKLTVAEAKAVCEFGGRLATYKLEAAARKTGFFHVCAAGWMAKGR 91  
Qy 112 VAYPTAFASQNCGGVGVGDYVPRPNKSEMDVFCYRMDKDVNC 155  
Db 92 VGYPIVPGNCGRGKGIIDYGIIRLNSERWDAYCYNPHAKEC 135

## RESULT 3

TSG-6 homolog PS4 - rabbit  
N:Alternate names: probable cell growth and differentiation protein, 12k; serum-inducibil  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 03-Mar-1994 #sequence\_revision 07-Jul-1995 #text\_change 08-Oct-1999  
C:Accession: A48055; A47290  
R:Feng, P.; Liaw, G.  
J. Biol. Chem. 268, 21453, 1993  
A:Reference number: A48055; MUID:94012707  
A:Accession: A48055  
A:Molecule type: mRNA  
A:Residues: 1-276 <FEN>  
A:Cross-references: GB:M86381; NID:g387866; PIDN:AAA03342.1; PID:g387867  
A:Note: this report replaces the sequence from reference A47290 and shares no similarity  
R:Feng, P.; Liaw, G.  
J. Biol. Chem. 268, 9387-9392, 1993  
A:Title: Identification of a novel serum and growth factor-inducible gene in vascular sm  
A:Reference number: A47290; MUID:93252803  
A:Accession: A47290

A:Molecule type: mRNA  
A:Residues: 1-276 <FEN>  
A:Cross-references: GB:M86381; NID:g387866; PIDN:AAA03342.1; PID:g387867  
A:Note: this report replaces the sequence from reference A47290 and shares no similarity  
R:Feng, P.; Liaw, G.  
J. Biol. Chem. 268, 9387-9392, 1993  
A:Title: Identification of a novel serum and growth factor-inducible gene in vascular sm  
A:Reference number: A47290; MUID:93252803  
A:Accession: A47290

Query Match 14.2% Score 264; DB 2; Length 276;  
Best Local Similarity 43.3% Pred. No. 3.1e-15;  
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

Qy 52 DTTGVFHLRSPGLQYKLTDFDKAREACANEATMATYNQLSYXOKAKYHLCSSAGWLETGR 111  
Db 32 EQAAGVYHREARSGYKLTVAEAKAVCEFGGRLATYKLEAAARKTGFFHVCAAGWMAKGR 91  
Qy 112 VAYPTAFASQNCGGVGVGDYVPRPNKSEMDVFCYRMDKDVNC 155  
Db 92 VGYPIVPGNCGRGKGIIDYGIIRLNSERWDAYCYNPHAKEC 135

## RESULT 4

hyaluronate-binding protein TSG-6 precursor - human  
A41735

C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 28-May-1999  
C:Accession: A41735; D53642  
R:Lee, T.H.; Wisniewski, H.G.; Vilcek, J.  
J. Cell Biol. 116, 545-557, 1992  
A:Title: A novel secretory tumor necrosis factor-inducible protein (TSG-6) is a membe  
A:Reference number: A41735; MUID:92112993  
A:Accession: A41735  
A:Molecule type: mRNA  
A:Residues: 1-277 <LEE>  
A:Cross-references: GB:M31165; NID:g339994; PIDN:AAB00792.1; PID:g1332377  
A:Experimental source: FS-4 fibroblast  
A:Note: sequence extracted from NCBI backbone (NCBIN:76828, NCBIIP:76833)  
R:Wisniewski, H.G.; Burgess, W.H.; Oppenheim, J.D.; Vilcek, J.  
Biochemistry 33, 7423-7429, 1994  
A:Title: TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable co  
A:Reference number: A53642; MUID:94271799  
A:Accession: D53642

A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 18-27 <WIS>  
C:Superfamily: C1r/C1s repeat homology; link protein repeat homology  
C:Keywords: glycoprotein

F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-277/Product: hyaluronate-binding protein TSG-6 #status predicted <MAT>  
F:53-128/Domain: link protein repeat homology <LNK>  
F:135-244/Domain: C1r/C1s repeat homology <C1R2>  
F:118,258/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.1% Score 262; DB 2; Length 277;  
Best Local Similarity 43.3% Pred. No. 4.7e-15;  
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

Qy 52 DTTGVFHLRSPGLQYKLTDFDKAREACANEATMATYNQLSYXOKAKYHLCSSAGWLETGR 111  
Db 32 ERAAGVYHREARSGYKLTVAEAKAVCEFGGRLATYKLEAAARKTGFFHVCAAGWMAKGR 91  
Qy 112 VAYPTAFASQNCGGVGVGDYVPRPNKSEMDVFCYRMDKDVNC 155  
Db 92 VGYPIVPGNCGRGKGIIDYGIIRLNSERWDAYCYNPHAKEC 135

## RESULT 5

I50421  
aggrecan precursor - chicken  
N:Alternate names: cartilage chondroitin sulfate proteoglycan core protein  
C:Species: Gallus gallus (chicken)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jan-2000  
C:Accession: I50421; S39796; S27356; A25442; A32002; I50216; A37072; B37072  
R:Li, H.; Schwartz, N.B.; Vertel, B.M.  
J. Biol. Chem. 268, 23504-23511, 1993  
A:Title: cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein  
A:Reference number: A48884; MUID:94043149  
A:Accession: I50421  
A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-2109 <LIX>  
A:Cross-references: GB:I21913; NID:g416133; PIDN:AAB19128.1; PID:g416134  
R:Chandrasekaran, L.; Tanzer, M.L.  
Biochem. J. 296, 885-887, 1993

A:Reference number: S39796; MUID:94107258  
A:Contents: annotation; erratum

A:Accession: S39796

A:Molecule type: mRNA

A:Residues: 1-361, 'DL', 364-600, 'R', 602-999, 'R', 1001-1028, 'P', 1030-1250, 'D', 1252-1602,  
A:Cross-references: GB:M88101

R:Chandrasekaran, L.; Tanzer, M.L.

Biochem. J. 288, 903-910, 1992

A:Title: Molecular cloning of chicken aggrecan. Structural analyses.

A:Reference number: S27356; MUID:93111968

A:Accession: S27356

A:Molecule type: mRNA

A:Residues: 1-361,'DL',364-600,'R',602-999,'R',1001-1028,'P',1030-1250,'D',1252-1549,'T'  
A:Cross-references: EMBL:M88101  
R:Sal, S.; Tanaka, T.; Koshier, R.A.; Tanzer, M.L.  
Proc. Natl. Acad. Sci. U.S.A. 83, 5081-5085, 1986  
A:Title: Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan  
A:Reference number: A25442; MUID:86259736  
A:Accession: A25442  
A:Molecule type: mRNA  
A:Residues: 1693-1795,'G',1797-1855,1894-2109 <SAI>  
A:Cross-references: GB:M13993; NID:g211654; PIDN:AAA48720.1; PID:g211655  
A:Experimental source: sternal cartilage  
R:Tanaka, T.; Har-El, R.; Tanzer, M.L.  
J. Biol. Chem. 263, 15831-15835, 1988  
A:Title: Partial structure of the gene for chicken cartilage proteoglycan core protein.  
A:Reference number: A32002; MUID:89008500  
A:Accession: A32002  
A:Molecule type: DNA  
A:Residues: 1893-1987,'S',1989-2022 <TAN>  
A:Note: the authors translated the codon TCC for residue 1787 as Phe  
R:Krueger, R.C.  
J. Biol. Chem. 265, 12088-12097, 1990  
A:Title: Chick cartilage chondroitin sulfate proteoglycan core protein: II. Nucleotide sequence  
A:Reference number: I50216; MUID:90307744  
A:Accession: I50216  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 'PA',1044-1559 <KRU>  
A:Cross-references: GB:M38187; NID:g211685; PIDN:AAA48731.1; PID:g555441  
R:Krueger Jr., R.C.; Fields, T.A.; Hildreth IV, J.; Schwartz, N.B.  
J. Biol. Chem. 265, 12075-12087, 1990  
A:Title: Chick cartilage chondroitin sulfate proteoglycan core protein. I. Generation and  
A:Reference number: A37072; MUID:90307743  
A:Accession: A37072  
A:Molecule type: protein  
A:Residues: 998-1015,'X',1017-1019,'X',1021-1023 <KR2>  
A:Note: amino end of 86K core peptide CS-A  
A:Accession: B37072  
A:Molecule type: protein  
A:Residues: 1247-1250,'D',1252-1272,'X',1274-1275 <KR3>  
A:Note: amino end of 75K core peptide CS-B  
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EC  
C:Keywords: alternative splicing  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-2109/Product: aggrecan #status predicted <MAT>  
F:44-131/Domain: immunoglobulin homology <IMM>  
F:166-243/Domain: link protein repeat homology <LNK1>  
F:264-346/Domain: link protein repeat homology <LNK2>  
F:537-614/Domain: link protein repeat homology <LNK3>  
F:635-716/Domain: link protein repeat homology <LNK4>  
F:1859-1890/Domain: EGF homology <EGF>  
F:1897-2017/Domain: C-type lectin homology <LCH>  
F:2024-2080/Domain: complement factor H repeat homology <FHD>  
Query Match 9.4%; Score 175; DB 1; Length 2109;  
Best Local Similarity 23.4%; Pred. No. 1.7e-06;  
Matches 75; Conservative 31; Mismatches 116; Indels 98; Gaps 12;  
QY 10 ECKSHYVGDGLNCEPEQLPIDRCLQDNCQCHA-----DAKCVDLHFQDTTVG 56  
DB 194 QCDAGLADQTVRYPIHILPRECRYGDKDFEPFGRVRYGVRETDETYDYCYAEQMQR--- 250  
QY 57 VFHLRSLPGQKLFDFKAREACANEATMATYNOLSYXQAKYHLCGAGWLETGRVAYPT 116  
DB 251 VFYATSP---BKFTFOEAFDKCHSLGARLATIGELYLAWKDMDCSAGWLADRSVYPI 307  
QY 117 APASQNCGVVGI--VDYGR-----PNKSEMMDVFCYRMKDYNV-----TXKVG-YV 162  
DB 308 SRARNPCGNLGVRTVYLPNANQGTGYPHPSRYDAICYSGDDFEALVPGLTFDEVGTET 367  
QY 163 GDGFSY-----SGNLLQVLMSPFSLNLF 186  
DB 368 GSAFTIQVTVQTEVELPLPRNVTEEEARGSTATLEPMEITATATLEYEAFTVLPDL--FA 425

QY 187 TEVIAYSNSSARGRAFLEHLTDL-----SIRGTLFXPQNSGLGE----- 225  
DB 426 TSVTVETASPRENTVTREITGIWAVPEEVTTSVSGTAF---TTGMAEVSVEEIAVTA 482  
QY 226 ---NETLSGRDIEHHLANVS 242  
DB 483 TPGLESASPFTIEDHLVQVT 502  
RESULT 6  
A28452  
proteoglycan core protein precursor, cartilage - rat  
N:Alternate names: aggrecan  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 21-Jan-2000  
C:Accession: A28452; A28453; A28454; A28455  
R:Doegge, K.; Sasaki, M.; Horiqan, E.; Hassell, J.R.; Yamada, Y.  
J. Biol. Chem. 262, 17757-17767, 1987  
A:Title: Complete primary structure of the rat cartilage proteoglycan core protein de  
A:Reference number: A28453; MUID:88087070  
A:Accession: A28453  
A:Molecule type: mRNA  
A:Residues: 1-2124 <DOE>  
R:Doegge, K.; Sasaki, M.; Horiqan, E.; Hassell, J.R.; Yamada, Y.  
J. Biol. Chem. 263, 10040a, 1988  
A:Reference number: A30069  
A:Contents: annotation; revision to residue 698  
R:Doegge, K.; Fernandez, P.; Hassell, J.R.; Sasaki, M.; Yamada, Y.  
J. Biol. Chem. 261, 8108-8111, 1986  
A:Title: Partial cDNA sequence encoding a globular domain at the C terminus of the ra  
A:Reference number: A28453; MUID:86250698  
A:Accession: A28453  
A:Molecule type: mRNA  
A:Residues: 1856-2124 <DO2>  
A:Cross-references: GB:M13518; NID:g206104; PIDN:AAA41836.1; PID:g206105  
R:Neame, P.J.; Christner, J.E.; Baker, J.R.  
J. Biol. Chem. 262, 17768-17778, 1987  
A:Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan amino-t  
A:Reference number: A28453; MUID:88087071  
A:Accession: A28453  
A:Molecule type: protein  
A:Residues: 20-37,'W',39-60,'E',62-64,'X',66-69;70-83;84,89-148,'L',150-238,'S',240,'  
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;  
C:Keywords: glycoprotein  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-2124/Product: proteoglycan core protein #status predicted <MAT>  
F:44-135/Domain: immunoglobulin homology <IMM>  
F:170-247/Domain: link protein repeat homology <LNK1>  
F:268-349/Domain: link protein repeat homology <LNK2>  
F:504-581/Domain: link protein repeat homology <LNK3>  
F:602-683/Domain: link protein repeat homology <LNK4>  
F:1914-2034/Domain: C-type lectin homology <LCH>  
F:2041-2097/Domain: complement factor H repeat homology <FHD>  
F:126.239,333,387,611,667,1842/Binding site: carbohydrate (Asn) (covalent) #status pr  
Query Match 9.2%; Score 171; DB 2; Length 2124;  
Best Local Similarity 28.8%; Pred. No. 3.7e-06;  
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;  
QY 57 VFHLRSLPGQKLFDFKAREACANEATMATYNOLSYXQAKYHLCGAGWLETGRVAYPT 116  
DB 488 VFHRYGSTRYSITFEAQOACIRTCAAIASPEQLQAAVEAGYEQCDAGWLDQTVRYPI 547  
QY 117 APASQNC---GSGVGVIVDYGPRPNKSEMDVFCYRMKDYNVCTXKVGXVGGFYSYGNLL 173  
DB 548 VSPRTPCVGDKDSPPGRTYGVRRPS-SETDYVCYVDK-----LEGEVF----- 590  
QY 174 QVLMSPFSLTNFLTEVLAYSNSSARGRAFLEHLTDLISRGTLFXPQNSGL 223  
DB 591 -----FATQMEQFTFOEAFCAQNAQT-LASTGQDIYAANSOGL 628

```
RESULT 7
T42630
aggreccan - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42630
R:Herling, T.M.; Kollar, J.; Huynh, T.D.
submitted to the EMBL Data Library, September 1996
A:Description: Complete coding sequence of bovine aggreccan: comparative structural analysis
A:Reference number: 222182
A:Accession: T42630
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2327 <HER>
A:Cross-references: EMBL:U76615; NID:g1730259; PID:g1730260; PIDN:AAB38524.1
A:Experimental source: articular chondrocytes
C:Superfamily: aggreccan; C-type lectin homology; complement factor H repeat homology; EGF
C:Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein

Query Match 9.2%; Score 170; DB 2; Length 2327;
Best Local Similarity 38.9%; Pred. No. 5e-06;
Matches 37; Conservative 12; Mismatches 42; Indels 4; Gaps 2;

QY 57 VFHLRSLPGVQKLFDRKAREACANEATMATYNOLSYXQKAKYHLCSSAGWLETGRVAYPT 116
Db 488 VFHVRPSSVSLTFEEAKQCLRTGALIASPEQLQAYEAGYEQCDAQWLDQGTVPRI 547

QY 117 AFASQNC---GSGVGVIVDYGPRPNKSEMDVFCY 148
Db 548 VSPRTPCVGDRKSSPGVRTYGVRP-PSETYDVICY 581

RESULT 8
LKRT2
proteoglycan link protein 2 precursor - rat
N:Alternate names: cartilage link protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-Aug-1986 #sequence_revision 24-Oct-1997 #text_change 13-Nov-1998
R:Rhodes, C.; Doege, K.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 263, 6063-6067, 1988
A:Title: Alternative splicing generates two different mRNA species for rat link protein.
A:Reference number: A28654; MUID:88198139
A:Accession: A28654
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-408 <RHO>
R:Doege, K.; Hassell, J.R.; Caterson, B.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 83, 3761-3765, 1986
A:Title: Link protein cDNA sequence reveals a tandemly repeated protein structure.
A:Reference number: A24880; MUID:86233314
A:Accession: A24880
A:Molecule type: mRNA
A:Residues: 180-408 <DOE>
R:Neame, P.J.; Christner, J.E.; Baker, J.R.
J. Biol. Chem. 261, 3519-3535, 1986
A:Title: The primary structure of link protein from rat chondrosarcoma proteoglycan aggregates
A:Reference number: A02869; MUID:86140139
A:Accession: A02869
A:Molecule type: protein
A:Residues: 16-33,'A',89-375,'W',377-408 <NEA>
C:Comment: This protein was extracted from rat chondrosarcoma.
C:Comment: Residues 259-277 and 358-376 (approximately) form epitopes for the species-specific
C:Comment: Link proteins interact with and stabilize aggregates of hyaluronic acid and
C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat
C:Keywords: cartilage; chondroitin sulfate proteoglycan; duplication; glycoprotein; hyal
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-408/Product: proteoglycan link protein 2 #status experimental <MAT>
F:108-195/Domain: immunoglobulin homology <IMM>
F:230-307/Domain: link protein repeat homology <LNK1>
F:328-404/Domain: link protein repeat homology <LNK2>

F:110/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:115-193,235-306,259-280,333-403,358-379/Disulfide bonds: #status experimental

Query Match 9.1%; Score 169.5; DB 1; Length 408;
Best Local Similarity 30.8%; Pred. No. 6.4e-07;
Matches 45; Conservative 20; Mismatches 52; Indels 29; Gaps 4;

QY 5 GKHKCECKSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCYDLHFQDITVGVFHLRSP 64
Db 189 GRYKCE-----VIEGL-----EDDTAVVALDLQGV---VFYFFPRL 221

QY 65 GOYKLTFRKAREACANEATMATYNOLSYXQKAKYHLCSSAGWLETGRVAYPTAFASQNC 124
Db 222 GRYNLNFHEARQAQLDQDAVIASFDQLYDAWRGGLDWCNAGWLSGDSVQYPIPKPREPCG 281

QY 125 --SGVVGVIVDYGPRPNKSEMDVFCY 148
Db 282 GQNTVPGVNRNYGFWDKSKRSYDVFCF 307

RESULT 9
S42938
proteoglycan link protein precursor - horse
N:Alternate names: cartilage link protein
C:Species: Equus caballus (domestic horse)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 22-Jun-1999
C:Accession: S42938
R:Dudhia, J.; Platt, D.
submitted to the EMBL Data Library, March 1994
A:Reference number: S42938
A:Accession: S42938
A:Molecule type: mRNA
A:Residues: 1-354 <DDD>
A:Cross-references: EMBL:X78077; NID:g459438; PIDN:CAA54987.1; PID:g459439
C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat
C:Keywords: cartilage; duplication; extracellular matrix; glycoprotein
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-354/Product: proteoglycan link protein #status predicted <MAT>
F:54-141/Domain: immunoglobulin homology <IMM>
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-350/Domain: link protein repeat homology <LNK2>
F:21,56/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:61-139/Disulfide bonds: #status predicted

Query Match 9.0%; Score 166.5; DB 1; Length 354;
Best Local Similarity 30.1%; Pred. No. 9.7e-07;
Matches 44; Conservative 22; Mismatches 51; Indels 29; Gaps 4;

QY 5 GKHKCECKSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCYDLHFQDITVGVFHLRSP 64
Db 135 GRYKCE-----VIEGL-----EDDTAVVALDLQGV---VFYFFPRL 167

QY 65 GOYKLTFRKAREACANEATMATYNOLSYXQKAKYHLCSSAGWLETGRVAYPTAFASQNC 124
Db 168 GRYNLNFHEARQAQLDQDAVIASFDQLYDAWRGGLDWCNAGWLSGDSVQYPIPKPREPCG 227

QY 125 --SGVVGVIVDYGPRPNKSEMDVFCY 148
Db 228 GQNTVPGVNRNYGFWDKSKRSYDVFCF 253

RESULT 10
LKCH
proteoglycan link protein precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 22-Jun-1999
C:Accession: A28305; A24881; A39097; B39097
R:Kiss, I.; Deak, F.; Mestric, S.; Dellus, H.; Soos, J.; Dekany, K.; Argraves, W.S.;
Proc. Natl. Acad. Sci. U.S.A. 84, 6399-6403, 1987
A:Title: Structure of the chicken link protein gene: exons correlate with the protein
A:Reference number: A28305; MUID:87317659
```



LKHU

proteoglycan link protein precursor - human  
N:Alternate names: cartilage link protein  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1991 #sequence.revision 31-Dec-1991 #text.change 22-Jun-1999  
C:Accession: S14914; S04244; S03868; A36308; S14926; S09309; S08041  
R:Dudhia, J.; Hardingham, T.E. 1990  
Nucleic Acids Res. 18, 1292, 1990  
A:Title: The primary structure of human cartilage link protein.  
A:Reference number: S14914; MUID: S0206798  
A:Accession: S14914  
A:Molecule type: mRNA  
A:Residues: 1-354 <DUD>  
A:Cross-references: EMBL:X17405; NID:q463246; PIDN:CAA35462.1; PID:q34378  
A:Note: the authors translated the codon GAT for residue 93 as Ala  
R:Dudhia, J.; Hardingham, T.E. 1989  
J. Mol. Biol. 206, 749-753, 1989  
A:Title: Appendix. Isolation and sequence of cDNA clones for pig and human cartilage lin  
A:Reference number: S04243  
A:Accession: S04244  
A:Molecule type: mRNA  
A:Residues: 223-354 <DUD2>  
A:Cross-references: EMBL:Y00166  
A:Note: the authors translated the codon CTG for residue 264 as Arg and TTG for residue  
R:Nguyen, Q.; Murphy, G.; Roughtley, P.J.; Mort, J.S.  
Biochem. J. 259, 61-67, 1989  
A:Title: Degradation of proteoglycan aggregate by a cartilage metalloproteinase. Evidenc  
A:Reference number: S03868; MUID:89246328  
A:Accession: S03868  
A:Molecule type: protein  
A:Residues: 16-35 <NGU>  
R:Osborne-Lawrence, S.L.; Sinclair, A.K.; Hicks, R.C.; Lacey, S.W.; Eddy Jr., R.L.; Byer  
Genomics 8, 562-567, 1990  
A:Title: Complete amino acid sequence of human cartilage link protein (CRTL1) deduced fr  
A:Reference number: A36308; MUID:91139126  
A:Accession: A36308  
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A:Molecule type: mRNA  
A:Residues: 1-354 <OSB>  
R:Dudhia, J.; Hardingham, T.E.  
Nucleic Acids Res. 18, 2214, 1990  
A:Title: The primary structure of human cartilage link protein.  
A:Reference number: S14926; MUID:90245703  
A:Accession: S14926  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-92,'A',94-354 <NUC>  
R:Nguyen, Q.; Mort, J.S.; Roughtley, P.J.  
Biochem. J. 266, 569-573, 1990  
A:Title: Cartilage proteoglycan aggregate is degraded more extensively by cathepsin L th  
A:Reference number: S09309; MUID:90197639  
A:Accession: S09309  
A:Molecule type: protein  
A:Residues: 16-38,'X',40-55,57-60,62-65,'X',67,'X',69-80 <NG2>  
C:Genetics:  
A:Gene: GDB:CRTL1  
A:Cross-references: GDB:125232; OMIM:115435  
A:Map position: 5q13-q14  
C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat  
C:Keywords: cartilage; duplication; extracellular matrix; glycoprotein  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-354/Product: proteoglycan link protein #status experimental <MAT>  
F:54-141/Domain: immunoglobulin homology <IMM>  
F:176-253/Domain: link protein repeat homology <LNK1>  
F:274-350/Domain: link protein repeat homology <LNK2>  
F:21,56/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:61-139/Disulfide bonds: #status predicted

Query Match 8.8%; Score 163.5; DB 1; Length 354;  
Best Local Similarity 30.1%; Pred. NO. 1.7e-06;  
Matches 44; Conservative 21; Mismatches 52; Indels 29; Gaps 4;

```

QY      5 GKHKCEKSHVYGDLNCEPEQLPIDRLQNQGCHADAKCVDLHFODTTGVGFHLRSLP 64
       : : |||          |   |   |   |   |   |   |   |   |   |   |
Db     135 GRVKCE-----VTGEL-----EDDTVVAALDQGV---VFPIFPRL 167

QY      65 GOYLKFLDKAREACANEAAATMATYNOLSYXOKAYHLCSAGWLETGRVAYPTAFASQCNG 124
       : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db     168 GRYNLNFHEAQACLDDQDAVIASFQQLDIWDNRGGLDWCNAGWLSDGSVOIPTKYPREPCG 227

QY      125 --SGVGIVDYGRPKNKSEMDVFCY 148
       : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db     228 GQNTVPGRNYGEFDKDSRYDVFCF 253


RESULT    13
S04243
proteoglycan link protein precursor - pig
N:Alternate names: cartilage link protein
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 22-Jun-1999
C:Accession: S04243; I47145
R:Dudhia, J.; Hardingham, T.E.
J. Mol. Biol. 206, 749-753, 1989
A:Title: Appendix. Isolation and sequence of cDNA clones for pig and human cartilage
A:Reference number: S04243
A:Accession: S04243
A:Molecule type: mRNA
A:Residues: 1-354 <DUD>
A:Cross-references: EMBL:Y00165; NID:g2009; PIDN:CAA68358.1; PID:g2010
R:Perkins, S.J.; Nealis, A.S.; Dudhia, J.; Hardingham, T.E.
J. Mol. Biol. 206, 737-753, 1989
A:Title: Immunoglobulin fold and tandem repeat structures in proteoglycan N-terminal
A:Reference number: I47145; MUID:89293937
A:Accession: I47145
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-354 <PER>
A:Cross-references: EMBL:Y00165; NID:g2009; PIDN:CAA68358.1; PID:g2010
C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeats
C:Keywords: cartilage; duplication; extracellular matrix; glycoprotein
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-354/Product: proteoglycan link protein #status predicted <NAT>
F:54-141/Domain: immunoglobulin homology <IMM>
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-350/Domain: link protein repeat homology <LNK2>
F:21,56/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:61-139/Disulfide bonds: #status predicted

```

Query Match 8.7%; Score 161.5; DB 1; Length 354;  
Best Local Similarity 29.5%; Pred. No. 2.6e-06;  
Matches 43; Conservative 22; Mismatches 52; Indels 29; Gaps 4;

QY	5	GKKCKESHVVGDLNCEPEQLPIDRCLQDNQGCHADAKCVDLHEQDTTGVGFHLRSPL	64
		::     ::  +   :	
Db	135	GRYKE-----VIEGL-----EDTAVVALDIKGV---VFYPFRL	167

```

QY      65  GQYKLTDFDKAREACANEATWATYNQLSYXQKAKYHLCAGWLETCRVAYPTAFASQNG 124
      1:| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      168  GRYNLNFHEAQOACLDODAVTASFQDLQYDAWRGGLDWCNAGWLSGSGVQYPIITKKPREPG 227

```

```
Qy 125 --SGVVGIVDYGPRPNKSEMWDVFCY 148
      : | | : | | : | | | :
Db 228 GQNTVPGVRN YGFWDKDKSR YDVFCE 253
```

RESULT  
A39086  
14

aggregan precursor - human  
N:Alternate names: chondroitin sulfate proteoglycan 1; large aggregating proteoglycan  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence.revision 10-Sep-1999 #text.change 28-Jan-2000  
C:Accession: A30086; S50206; A43919; S46659; S66389; S68646; B43919; C43919  
R:Doegi, K.J.; Sakaki, M.; Kimura, T.; Yamada, Y.

```
Search completed: April 4, 2001, 13:04:41
Job time: 44 sec
```

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:03:56 ; Search time 13.89 Seconds  
(without alignments)  
456.360 Million cell updates/sec

Title: US-09-466-778-11  
Perfect score: 1857  
Sequence: 1 MTGPGKHCKECSHYVGDGL.....ALAAYSYFINKTIGFXHF 353

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/PTUS.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	262	14.1	277	1	US-08-024-868-2
2	262	14.1	277	2	US-08-242-097-2
3	262	14.1	277	4	PCT-US96-11995-1
4	167	9.0	326	1	US-08-225-477B-6
5	167	9.0	326	2	PCT-US95-04353-6
6	160	8.6	98	2	US-08-242-097-4
7	157.5	8.5	371	1	US-08-225-477B-8
8	157.5	8.5	371	4	PCT-US95-04353-8
9	157.5	8.5	1257	1	US-08-340-428B-49
10	157	8.5	329	1	US-08-225-477B-3
11	157	8.5	329	4	PCT-US95-04353-3
12	156	8.4	333	1	US-08-225-477B-4
13	156	8.4	333	4	PCT-US95-04353-4
14	155.5	8.4	912	4	PCT-US95-03747-2
15	155.5	8.3	97	2	US-08-242-097-5
16	155.5	8.1	328	1	US-08-225-477B-5
17	150.5	8.1	328	4	PCT-US95-04353-5
18	150.5	8.1	2409	5	5180808-2
19	148	8.0	908	4	PCT-US95-03747-3
20	147	7.9	378	1	US-08-225-477B-9
21	147	7.9	378	4	PCT-US95-04353-9
22	107	5.8	322	2	US-08-892-880-2
23	105	5.7	810	2	US-08-820-170A-34
24	105	5.7	810	3	US-09-055-699-34
25	105	5.7	810	3	US-09-273-565-34
26	103.5	5.6	717	3	US-08-872-855-9
27	99	5.3	735	3	US-09-191-647-9
28	98.5	5.3	363	1	US-07-946-497-7

29	98.5	5.3	363	1	US-08-483-322-7
30	98.5	5.3	363	2	US-08-478-882-7
31	97.5	5.3	339	2	US-08-892-880-3
32	97.5	5.3	503	1	US-07-946-497-2
33	97.5	5.3	503	2	US-08-483-322-2
34	97.5	5.3	503	2	US-08-478-882-2
35	97.5	5.3	1480	3	US-09-191-647-7
36	97.5	5.3	1480	4	PCT-US91-09055-2
37	97	5.2	729	3	US-08-872-855-8
38	96.5	5.2	721	3	US-08-872-855-7
39	96	5.2	361	1	US-07-946-497-6
40	96	5.2	361	2	US-08-483-322-6
41	96	5.2	361	2	US-08-478-882-6
42	94.5	5.1	720	3	US-08-872-855-4
43	92.5	5.0	362	5	5504194-2
44	91.5	4.9	1404	2	US-08-400-159-2
45	91.5	4.9	1404	3	US-08-611-729A-2

ALIGNMENTS

RESULT 1  
US-08-024-868-2  
; Sequence 2, Application US/08024868  
; Patent No. 5386013  
; GENERAL INFORMATION:  
; APPLICANT: Lee, Tae Ho  
; APPLICANT: Wisniewski, Hans Georg  
; APPLICANT: Vilcek, Jan  
; TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA  
; CODING THEREOF AND USES THEREOF  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/024,868  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/642,312  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Livnat, Shmuel  
REGISTRATION NUMBER: 33,949  
REFERENCE/DOCKET NUMBER: VILCEK=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-628-5197  
TELEFAX: 212-737-3528  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 277 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-024-868-2

Query Match 14.1%; Score 262; DB 1; Length 277;  
Best Local Similarity 43.3%; Pred. No. 8e-21; Indels 0; Gaps 0;  
Matches 45; Conservative 16; Mismatches 43;

Db 32 ERAAGVYHREARSKYKLTAEAKAVCEFGGHLATYKQLEARKIGFHVCAAGWMAKGR 91  
QY 112 VAYPTAFASQNGSGVGVVDYGRPNKSEMMDVFCYRMKDVC 155  
Db 92 VGYPIVPGNCGFGTKGIIDYGRNLRSERWDAYCYNPHAKEC 135

## RESULT 2

US-08-242-097-2  
; Sequence 2, Application US/08242097  
; Patent No. 5846763  
; GENERAL INFORMATION:  
; APPLICANT: Lee, Tae Ho  
; APPLICANT: Wisniewski, Hans Georg  
; APPLICANT: Vilcek, Jan  
; TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding  
; TITLE OF INVENTION: Thereof and Uses Thereof  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/242,097  
; FILING DATE: 13-MAY-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: LEE26/VILCEK-1B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-628-5197  
; TELEFAX: 212-737-3528  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 277 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-242-097-2

Query Match 14.1%; Score 262; DB 2; Length 277;  
Best Local Similarity 43.3%; Pred. No. 8e-21;  
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;  
QY 52 DTTVGVFHLRSLPGQYKLTFDKAREACANEATMATYNQLSYXQKAKYHLCISAGWLETGR 111  
Db 32 ERAAGVYHREARSKYKLTAEAKAVCEFGGHLATYKQLEARKIGFHVCAAGWMAKGR 91  
QY 112 VAYPTAFASQNGSGVGVVDYGRPNKSEMMDVFCYRMKDVC 155  
Db 92 VGYPIVPGNCGFGTKGIIDYGRNLRSERWDAYCYNPHAKEC 135

## RESULT 3

PCT-US96-11995-1  
; Sequence 1, Application PC/TUS9611995  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: A Pharmaceutical Composition Containing TSG-6  
; TITLE OF INVENTION: Protein for Treating Inflammatory Diseases and Cancer-Related  
; TITLE OF INVENTION: Methods of Using same  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/11995  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/001,311  
; FILING DATE: 20-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: WISNIEWSKI-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-628-5197  
; TELEFAX: 212-737-3528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 277 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US96-11995-1

Query Match 14.1%; Score 262; DB 4; Length 277;  
Best Local Similarity 43.3%; Pred. No. 8e-21;  
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;  
QY 52 DTTVGVFHLRSLPGQYKLTFDKAREACANEATMATYNQLSYXQKAKYHLCISAGWLETGR 111  
Db 32 ERAAGVYHREARSKYKLTAEAKAVCEFGGHLATYKQLEARKIGFHVCAAGWMAKGR 91  
QY 112 VAYPTAFASQNGSGVGVVDYGRPNKSEMMDVFCYRMKDVC 155  
Db 92 VGYPIVPGNCGFGTKGIIDYGRNLRSERWDAYCYNPHAKEC 135

## RESULT 4

US-08-225-477B-6  
; Sequence 6, Application US/08225477B  
; Patent No. 5635370  
; GENERAL INFORMATION:  
; APPLICANT: Susan Hockfield  
; APPLICANT: Diane M. Jaworski  
; TITLE OF INVENTION: BEHAB, A Brain Hya-  
; TITLE OF INVENTION: Iuronan-Binding Protein  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: St. Onge Steward Johnston & Reens  
; STREET: 986 Bedford Street  
; CITY: Stamford  
; STATE: CT  
; COUNTRY: United States  
; ZIP: 06905  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" 1.44 Mb diskette  
; COMPUTER: IBM PC  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: Word Processor  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/225,477B  
; FILING DATE: April 8, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mary M. Krinsky

REGISTRATION NUMBER: 32423  
REFERENCE/DOCKET NUMBER: 1751-P0004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 203-324-6155  
TELEFAX: 203-327-1096  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 326 residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
DESCRIPTION: functional domains  
FRAGMENT TYPE:  
FEATURE:  
NAME/KEY: rat link protein  
PUBLICATION INFORMATION:  
AUTHORS: Doege, K., Hassell, J.R., Ca-  
AUTHORS: tersen, B., and Yamada, Y.  
TITLE: Link protein cDNA sequence reveals a  
TITLE: tandemly repeated protein sequence.  
JOURNAL: Proc. Natl. Acad. Sci. USA  
VOLUME: 83  
PAGES: 3761-3765  
DATE: 1986  
US-08-225-477B-6

Query Match 9.0%; Score 167; DB 1; Length 326;  
Best Local Similarity 31.5%; Pred. No. 3.2e-10;  
Matches 46; Conservative 21; Mismatches 49; Indels 30; Gaps 5;

QY 5 GKHKCEKSHYVGDLNCEPEQLPIDRCLQDNCQCHADAKCVDLHFQDPTVGVFHLRSP 64  
DB 109 GRYKCE-----VIEGL-----EDDTAVVALELQGV---VFPYFPRL 141  
QY 65 GOYKLTDFKAREACANEATMATYNOLSYXOKAKYHLCGAGWLETGRVAYPTAFASONG 124  
DB 142 GRYNLNPFHARQAQLDQDAVIAFDQLYDAWRGLDWCNAGWLSGDSGVQYPTITKPRECG 201  
QY 125 --SGVGVGDYGRPNKSEMDVFCY 148  
DB 202 GQNTVPGVRNYG-FWDKDSRYDVFCF 226

RESULT 5  
PCT-US95-04353-6  
Sequence 6, Application PC/TUS9504353  
GENERAL INFORMATION:  
APPLICANT: Susan Hockfield  
APPLICANT: Diane M. Jaworski  
TITLE OF INVENTION: BEHAB, A Brain Hya-  
TITLE OF INVENTION: Luronan-Binding Protein  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: St. Onge Steward Johnston & Reens  
STREET: 986 Bedford Street  
CITY: Stamford  
STATE: CT  
COUNTRY: United States  
ZIP: 06905  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" 1.44 Mb diskette  
OPERATING SYSTEM: MS DOS  
SOFTWARE: Word Processor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04353  
FILING DATE:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/225,477  
FILING DATE: April 8, 1994  
ATTORNEY/AGENT INFORMATION:

NAME: Mary M. Krinsky  
REGISTRATION NUMBER: 32423  
REFERENCE/DOCKET NUMBER: 1751-P0004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 203-324-6155  
TELEFAX: 203-327-1096  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 326 residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
DESCRIPTION: functional domains  
FRAGMENT TYPE:  
FEATURE:  
NAME/KEY: rat link protein  
PUBLICATION INFORMATION:  
AUTHORS: Doege, K., Hassell, J.R., Ca-  
AUTHORS: tersen, B., and Yamada, Y.  
TITLE: Link protein cDNA sequence reveals a  
TITLE: tandemly repeated protein sequence.  
JOURNAL: Proc. Natl. Acad. Sci. USA  
VOLUME: 83  
PAGES: 3761-3765  
DATE: 1986  
PCT-US95-04353-6

Query Match 9.0%; Score 167; DB 4; Length 326;  
Best Local Similarity 31.5%; Pred. No. 3.2e-10;  
Matches 46; Conservative 21; Mismatches 49; Indels 30; Gaps 5;

QY 5 GKHKCEKSHYVGDLNCEPEQLPIDRCLQDNCQCHADAKCVDLHFQDPTVGVFHLRSP 64  
DB 109 GRYKCE-----VIEGL-----EDDTAVVALELQGV---VFPYFPRL 141  
QY 65 GOYKLTDFKAREACANEATMATYNOLSYXOKAKYHLCGAGWLETGRVAYPTAFASONG 124  
DB 142 GRYNLNPFHARQAQLDQDAVIAFDQLYDAWRGLDWCNAGWLSGDSGVQYPTITKPRECG 201  
QY 125 --SGVGVGDYGRPNKSEMDVFCY 148  
DB 202 GQNTVPGVRNYG-FWDKDSRYDVFCF 226

RESULT 6  
US-08-242-097-4  
Sequence 4, Application US/08242097  
Patent No. 5846763  
GENERAL INFORMATION:  
APPLICANT: Lee, Tae Ho  
APPLICANT: Wisniewski, Hans Georg  
APPLICANT: Vlicek, Jan  
TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding  
TITLE OF INVENTION: Therefor and Uses Thereof  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PC-DOS/MS-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/242,097  
FILING DATE: 13-MAY-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:



Query Match 8.5%; Score 157.5; DB 4; Length 371;  
Best Local Similarity 27.9%; Pred. No. 4.3e-09;  
Matches 46; Conservative 22; Mismatches 60; Indels 37; Gaps 5;

QY 10 ECKSHVVDGLNCEPEQLPIDRCLQD-----NGQCHADAKCVDLHF 50  
DB 201 QCDAGWLSDDTVRYPIQNPRACYGDMGYPGVNRNYGVVGPDDLVDYCYAE-----DLN- 255  
QY 51 QDVTGVFHLRSPGLQYKLTDFDKAREACANEATMATYNQLSYXOKAKYHLCAGWLETG 110  
DB 256 -----GELFLGAPPG--KLTWEARDYCLERGAQTASTGQLYAAWNGGLDRCSPGWLADG 308  
QY 111 RVAYPTAFASQNGSGVGIYDYGPRPN-----KSEMDVFCYR 149  
DB 309 SVRYPIIPTSQRGGGLPGVTLFLFPNQTFPPSKONRFNYICFR 353

RESULT 9  
US-08-340-428B-49  
; Sequence 49, Application US/08340428B  
; Patent No. 5648465  
; GENERAL INFORMATION:  
; APPLICANT: MARGOLIS, Richard U.  
; APPLICANT: RAUCH, Uwe  
; APPLICANT: MARGOLIS, Renee K.  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A  
; TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/340.428B  
FILING DATE: 14 No. 5648465ember 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/922,911  
FILING DATE: 03 August 1992  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: Margolis-1A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-340-428B-49

Query Match 8.5%; Score 157.5; DB 1; Length 1257;  
Best Local Similarity 34.2%; Pred. No. 2.8e-08;  
Matches 41; Conservative 19; Mismatches 55; Indels 7; Gaps 3;  
QY 47 DLHFQDVTGVFHLRSPGLQYKLTDFDKAREACANEATMATYNQLSYXOKAKYHLCAGW 106  
DB 150 DLVTELVGVFHYRAARDRYALTFAEAOEACHLSSATIAAPRHQAQAFEDGFDCNDAGW 209

QY 107 LETGRVAYPTAFASQNG---GSGVVGIVDYGPRPNKSEMDVFCYRMKDVNCTKXGVYVG 163  
DB 210 LSDRTVRYPIITQSRPCGYGDRSSLPVRSYG--RRDPQELIDYVCFARE---LGGEVFTVG 265

RESULT 10  
US-08-225-477B-3  
; Sequence 3, Application US/08225477B  
; Patent No. 5635370  
; GENERAL INFORMATION:  
; APPLICANT: Susan Hockfield  
; APPLICANT: Diane M. Jaworski  
; TITLE OF INVENTION: BEHAB, A Brain Hya-  
; TITLE OF INVENTION: luronan-Binding Protein  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: St. Onge Steward Johnston & Reens  
; STREET: 986 Bedford Street  
; CITY: Stamford  
; STATE: CT  
; COUNTRY: United States  
; ZIP: 06905  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" 1.44 Mb diskette  
COMPUTER: IBM PC  
OPERATING SYSTEM: MS DOS  
SOFTWARE: Word Processor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/225.477B  
FILING DATE: April 8, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary M. Krinsky  
REGISTRATION NUMBER: 32423  
REFERENCE/DOCKET NUMBER: 1751-P0004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 203-324-6155  
TELEFAX: 203-327-1096  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 329 residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
; MOLECULE TYPE:  
DESCRIPTION: polypeptide  
FRAGMENT TYPE: functional domains  
; FEATURE:  
NAME/KEY: rat aggregan  
PUBLICATION INFORMATION:  
AUTHORS: Doege, K., Sasaki, M., Hori-  
AUTHORS: gan, E., Hassell, J.R., and Yamada, Y.  
TITLE: Complete primary structure of the rat  
TITLE: cartilage proteoglycan core protein deduced from  
TITLE: cDNA clones.  
JOURNAL: J. Biol. Chem.  
VOLUME: 262  
PAGES: 17757-17767  
DATE: 1987  
US-08-225-477B-3

Query Match 8.5%; Score 157; DB 1; Length 329;  
Best Local Similarity 26.8%; Pred. No. 4.1e-09;  
Matches 44; Conservative 23; Mismatches 67; Indels 30; Gaps 6;

QY 5 GKHKCEKSHYVGDLNCEPEQLPIDRC-----LDNGQCHADAKCVDLHF 50  
DB 175 GFHQCD--AGWLADQVRYPIHTPREGCYGDKDFPGVYTGIRDTNETY-DVYC----F 227  
QY 51 QDVTGVFHLRSPGLQYKLTDFDKAREACANEATMATYNQLSYXOKAKYHLCAGWLETG 110  
DB 228 AEEMGEFEVATSP---EKFTFOEAAANE CRTVGARLATTGQLYLAWQGGMDMCSAGWLADR 284





Search completed: April 4, 2001, 13:04:18  
Job time: 22 sec



Wed Apr 4 13:03:52 2001

us-09-466-778-11.rspt

SEQ ID 11, Database: SPTREMBL-15  
AC NOs.: Q9NRY3, Q9UF98

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:04:01; Search time 27.49 Seconds

(without alignments)

1505.071 Million cell updates/sec

Title: US-09-466-778-11

Perfect score: 1857

Sequence: 1 MTGPGKHCKESHVVDGL.....ALAAVSFRINRKTIGFXHF 353

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB-seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL.15.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organalle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1757	94.6	897	Q9NRY3	Q9NRY3 homo sapien
2	1468	79.1	1069	Q9UF98	Q9UF98 homo sapien
3	662	35.6	2212	Q93072	Q93072 homo sapien
4	645.5	34.8	2570	Q9NRY15	Q9NRY15 homo sapien
5	264	14.2	275	Q08859	Q08859 mus musculus
6	177	9.5	2109	Q97987	Q97987 gallus gall
7	169.5	9.1	355	Q921X7	Q921X7 mus musculus
8	162	8.7	1321	Q14594	Q14594 homo sapien
9	161.5	8.7	655	Q088564	Q088564 rattus norv
10	161.5	8.7	1290	Q9W6E1	Q9W6E1 gallus gall
11	160.5	8.6	2394	Q07610	Q07610 bos taurus
12	153	8.2	192	Q02817	Q02817 oryctolagus
13	153	8.2	656	Q07612	Q07612 bos taurus
14	151.5	8.2	1643	Q07611	Q07611 bos taurus
15	151.5	8.2	3381	Q07609	Q07609 bos taurus
16	126.5	6.8	103	Q9TTB3	Q9TTB3 sus scrofa
17	120.5	6.5	103	Q046380	Q046380 oryctolagus
18	110	5.9	396	Q9W6S4	Q9W6S4 gallus gall
19	109	5.9	302	Q9UNF4	Q9UNF4 homo sapien

20	109	5.9	322	4	Q9Y5Y7	Q9Y5Y7 homo sapien
21	105	5.7	1584	5	Q93791	Q93791 caenorhabdi
22	104	5.6	3507	5	Q93587	Q93587 caenorhabdi
23	103.5	5.6	717	13	P87357	P87357 brachydanio
24	102.5	5.5	778	13	Q91BGA	Q91BGA xenopus lae
25	100	5.4	816	11	Q70474	Q70474 rattus norv
26	99.5	5.4	682	4	Q9NTM2	Q9NTM2 homo sapien
27	99.5	5.4	728	13	Q90656	Q90656 gallus gall
28	99.5	5.4	802	13	Q57462	Q57462 brachydanio
29	99.5	5.4	3623	4	Q60494	Q60494 homo sapien
30	99	5.3	780	11	Q9QYX8	Q9QYX8 mus musculu
31	99	5.3	1440	5	Q20204	Q20204 caenorhabdi
32	99	5.3	2906	11	Q9WUH9	Q9WUH9 rattus norv
33	98	5.3	294	4	Q92493	Q92493 homo sapien
34	98	5.3	2531	5	Q16004	Q16004 lytechinus
35	97.5	5.3	780	11	Q08779	Q08779 rattus norv
36	97.5	5.3	1480	5	Q9V7F8	Q9V7F8 drosophila
37	97.5	5.3	1504	5	Q9XYV4	Q9XYV4 drosophila
38	97.5	5.3	1504	5	Q9V7F9	Q9V7F9 drosophila
39	96	5.2	364	11	Q70509	Q70509 rattus norv
40	96	5.2	742	4	Q9UJ36	Q9UJ36 homo sapien
41	95.5	5.1	364	6	Q97569	Q97569 ceratotheri
42	95.5	5.1	721	13	Q91902	Q91902 xenopus lae
43	95	5.1	261	10	Q24530	Q24530 vitis vinif
44	94.5	5.1	1328	13	P79754	P79754 fugu rubrip
45	94	5.1	637	10	Q9SF50	Q9SF50 arabidopsis

## ALIGNMENTS

RESULT 1  
Q9NRY3 PRELIMINARY; PRT; 897 AA.

AC Q9NRY3; 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE CD44-LIKE PRECURSOR FELL;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tao Q., Zhang W., Cao X.;  
RT "Molecular cloning and characterization of human FELL sharing homology with CD44."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF160476; AAF82398.1;  
SQ SEQUENCE 897 AA; 97585 MW; EB920AF36101E388 CRC64;

Query Match 94.6%; Score 1757; DB 4; Length 897;  
Best Local Similarity 93.8%; Pred. NO. 8.6e-162;  
Matches 331; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

QY 1 MTGPGKHCKESHVVDGLNCEPEQLPDRCLDNGOCHADAKCVDLHFQDFTVGVFHL 60  
DB 489 MTGPGKHCKESHVVDGLNCEPEQLPDRCLDNGOCHADAKCVDLHFQDFTVGVFHL 548  
QY 61 RSLPGYKLTFTDKARACANEATMATYNQLSYXQAKYHLCSAGWLTGRVAYPTAFAS 120  
DB 549 RSLPGYKLTFTDKARACANEATMATYNQLSYXQAKYHLCSAGWLTGRVAYPTAFAS 608  
QY 121 QNCGSGWGVIVDGPDPNPKSEMDVFCYRMKDVNCTXXVGVVGDGCFYSYGNLLQVLSFP 180  
DB 609 QNCGSGWGVIVDGPDPNPKSEMDVFCYRMKDVNCTXXVGVVGDGCFYSYGNLLQVLSFP 668  
QY 181 SLTNFTLTVLAYSNSARGAFLEHLTDLSTIRGTLFXPONSGLGENETLSGRDIEHLAN 240  
DB 669 SLTNFTLTVLAYSNSARGAFLEHLTDLSTIRGTLFXPONSGLGENETLSGRDIEHLAN 728



```
QY 241 VSMFFNDLVNGTTTQTQLGSKLLITDRQDPLHPTETRCVDRDTEWIDICASNGITHVI 300
DB 729 VSMFFNDLVNGTTTQTQLGSKLLITASQDPLQPTETRFVDGRAILQWIFASNGIHHVI 788
QY 301 SRXLKAPAPVTLHTXILGXGIFXXIILVTGAVALAAYSFRINRRTIGFXXHF 353
DB 789 SRPLKAPAPVTLHTXILGXGIFXXIILVTGAVALAAYSFRINRRTIGFQHF 841

RESULT 2
Q9UF98 PRELIMINARY; PRT; 1069 AA.
AC Q9UF98;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHETICAL 115.7 KDA PROTEIN (FRAGMENT).
GN DKF2P434E0321.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL133021; CAB61358.1;
DR HSP: P98066; 1TSG.
DR INTERPRO: IPR000538;
DR INTERPRO: IPR000561;
DR INTERPRO: IPR000782;
DR PFAM: PF00008; EGF; 8.
DR PFAM: PF00193; Xlink; 1.
DR PRINTS: PR00193; Xlink; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_5.
DR PROSITE: PS01186; EGF_2; 13.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 2.
DR KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1069 AA; 115666 MW; 04B0950932164D63 CRC64;
```

```
Query Match 79.18; Score 1468; DB 4; Length 1069;
Best Local Similarity 80.58; Pred. No. 1.2e-133;
Matches 284; Conservative 4; Mismatches 29; Indels 36; Gaps 2;

QY 1 MTGPGKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL 60
DB 697 MTGPGKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL 756
QY 61 RSPGQYKLTDFDKAREACANAATMYNQLSYQAKYHLCASGWLGTGRVAYPTAFAS 120
DB 757 RSPGQYKLTDFDKAREACANAATMYNQLSYQAKYHLCASGWLGTGRVAYPTAFAS 816
QY 121 QNCGSGVGVIVDYGPRPNKSEMDVFCYRMKDVCNCTKKVGVGDGFSYSGNLLQVLMSEF 180
DB 817 QNCGSGVGVIVDYGPRPNKSEMDVFCYRMKG -----SAGLFQQLSSRP 860
QY 181 SUTNLTFLVAYSNSARGRAFLHSLTDLIRGTLPQNSGLGENETLSGRDIEHHLA 240
DB 861 CIS-----RTPDLSIRGTLFYPQNSGLGENETLSGRDIEHHLA 900
QY 241 VSMFFNDLVNGTTTQTQLGSKLLITDRQDPLHPTETRCVDRDTEWIDICASNGITHVI 300
DB 901 VSMFFNDLVNGTTTQTQLGSKLLITASQDPLQPTETRFVDGRAILQWIFASNGIHHVI 960
QY 301 SRXLKAPAPVTLHTXILGXGIFXXIILVTGAVALAAYSFRINRRTIGFXXHF 353
DB 961 SRPLKAPAPVTLHTXILGXGIFXXIILVTGAVALAAYSFRINRRTIGFQHF 1013
```

```
RESULT 3
Q93072 PRELIMINARY; PRT; 2212 AA.
AC Q93072;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE MYELOBLAST KIAA0246 PROTEIN (FRAGMENT).
GN KIAA0246.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=9719154; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.
RL DNA Res. 3:321-329(1996).
DR EMBL: D87433; BAA1337.1;
DR HSP: P98066; 1TSG.
DR INTERPRO: IPR000538;
DR INTERPRO: IPR000561;
DR INTERPRO: IPR000782;
DR INTERPRO: IPR001128;
DR PFAM: PF00008; EGF; 13.
DR PFAM: PF00193; Xlink; 1.
DR PRINTS: PR01265; LINKMODULE.
DR PROSITE: PS00022; EGF_1; UNKNOWN_5.
DR PROSITE: PS01186; EGF_2; 13.
DR PROSITE: PS01241; LINK; UNKNOWN_1.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 2.
DR KW Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 2212 AA; 237451 MW; 4A95A60504129134 CRC64;
```

```
Query Match 35.68; Score 662; DB 4; Length 2212;
Best Local Similarity 40.28; Pred. No. 4.9e-55;
Matches 143; Conservative 47; Mismatches 158; Indels 8; Gaps 5;

QY 2 TGPGRKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL 60
DB 1796 TGLNTRRCECHAGYVGDGLQCLESEPPVDRCLGQPPPCHEADACTDLHFQEKRAVGFHL 1855
QY 61 RSPGQYKLTDFDKAREACANAATMYNQLSYQAKYHLCASGWLGTGRVAYPTAFAS 120
DB 1856 QATSGYGLNFSERAEACACQAGVLAAPPQLSAQQLGFHLCLMELANGSTAHPPVFPV 1915
QY 121 QNCGSGVGVIVDYGPRPNKSEMDVFCYRMKDVCNCTKKVGVGDGFSYSGNLLQVLMSEF 179
DB 1916 ADCGNGRGGVIVSGARKNLSEWDAYCFRQDVACRCRNGFVGDSICNGKLLDLAAT 1975
QY 180 PSUTNLTFLVAYSNSARGRAFLHSLTDLIRGTLPQNSGLGENETLSGRDIEHHLA 239
DB 1976 ANFTSYGMLLYANATQKGLDFLDDELTYKTLFVPVNEGFPVDMTUSGPDLEIHLAS 2035
QY 240 NVSMFFNDLVNGTTTQTQLGSKLLITDR---QDPLHPTETRCVDRDTEWIDICASNGI 296
DB 2036 NATLLSAN-ASQGLLPAHSGLSLLIISDAGPDNNSWAPAGTVVVSRIIVMDIAFNGL 2094
QY 297 THVISRLKAPAPVTLHTXILGXGIFXXIILVTGAV--ALAAYSFRINRRTIGF 350
DB 2095 IHALASPLLAPPQPAVLAPAPPAAGVAVLAAGALIGLVAGALYLRARGKPMGF 2150
```



File Copy

Wed Apr 4 13:03:50 2001

us-09-466-778-11.rag

SEQ ID 11, Database: A-Geneseq-36

Page 2

Ac NO. W84087

PR 14-JAN-1991; 91US-0642312.  
XX (UUNY ) UNIV NEW YORK STATE.  
XX Lee TH, Vilcek J, Wisniewski HG;  
XX WPI; 1990-284330/34.  
XX N-PSDB; Q27190.  
XX  
XX Tumour necrosis factor-induced (glyco)protein mol. TSG-6 and its  
XX antibody - for treatment and diagnosis of chronic inflammatory  
XX conditions e.g. rheumatoid arthritis, infections, sepsis and  
XX cancer  
XX  
XX Claim 1; Fig 3; 122pp; English.  
XX  
XX The sequence is that of tumour necrosis factor (TNF) induced  
XX glycoprotein TSG-6. TSG-6 is induced in connective tissue by TNF  
XX and interleukin-7 and interferes with cell adhesion to their cells or  
XX the extracellular matrix. It and antibodies specific for it are useful  
XX in diagnosis (including identifying susceptibility) and treatment of  
XX various disorders associated with cytokine (in)activation e.g. chronic  
XX inflammation (esp. where associated with proteoglycan breakdown, such  
XX as rheumatoid arthritis) cancer (esp. development of metastases) and  
XX infections (esp. those caused by Gram negative bacteria).  
XX  
XX Sequence 277 AA;  
XX  
XX  
XX Query Match 14.1%; Score 262; DB 13; Length 277;  
XX Best Local Similarity 43.3%; Pred. No. 1.6e-19;  
XX Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;  
XX  
XX 52 DTTGVFHLRSPGLGQYKLFDFKAREACANEATMATYNOLSYOKAKYHLCGAGWLETR 111  
XX 32 eraagvyhrearsgkyklytaeakvcefgghlatykglearkigfhvcaagwmakgr 91  
XX  
XX 112 VAYPTAFASQNCGSGVGVIVDYGPENKSEMMDVFCYRMDVNC 155  
XX 92 vgyplvpgncpgcgtgldygrlnrserwdaycynphakec 135  
XX  
XX  
XX RESULT 2  
XX W13654  
XX ID W13654 standard; protein; 277 AA.  
XX AC W13654;  
XX  
XX 07-NOV-1997 (first entry)  
XX  
XX Tumour necrosis factor-stimulated gene-6 protein.  
XX  
XX TSG-6; human; tumour necrosis factor-stimulated gene-6 protein; treat;  
XX inflammatory disease; cancer-related pathology; inhibit; metastasis;  
XX invasive tumour growth; rheumatoid arthritis; cachexia; atherosclerosis;  
XX systemic lupus erythematosus; rational drug design.  
XX  
XX Homo sapiens.  
XX  
XX WO9704075-A1.  
XX  
XX 06-FEB-1997.  
XX  
XX 19-JUL-1996; 96WO-US11995.  
XX  
XX 20-JUL-1995; 95US-0001311.  
XX  
XX (UUNY ) UNIV NEW YORK STATE.  
XX  
XX Cronstein BN, Vilcek J, Wisniewski H;  
XX WPI; 1997#132619/12.  
XX  
XX

XX New TSG-6 compositions to treat inflammatory diseases and cancer -  
XX can be complexed with inter-alpha-inhibitor, to treat auto-immune  
XX disease, cancer-related pathology etc. can also be administered as  
XX TSG-6 DNA  
XX  
XX Claim 2; Page 58-59; 76pp; English.  
XX  
XX This protein is the human tumour necrosis factor-stimulated gene-6  
XX protein (TSG-6). It can be used in a claimed pharmaceutical composition  
XX for treating inflammatory diseases and disorders, or cancer-related  
XX pathologies. The composition or DNA encoding TSG-6 may be used to inhibit  
XX tumour metastases, invasive tumour growth or cancer-related pathologies.  
XX They may also be used in the treatment of inflammatory conditions  
XX including rheumatoid arthritis, systemic lupus erythematosus, and  
XX infections such as cachexia, atherosclerosis etc. TSG-6 can also be used  
XX to design drugs to treat these disorders.  
XX  
XX Sequence 277 AA;  
XX  
XX Query Match 14.1%; Score 262; DB 18; Length 277;  
XX Best Local Similarity 43.3%; Pred. No. 1.6e-19;  
XX Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;  
XX  
XX 52 DTTGVFHLRSPGLGQYKLFDFKAREACANEATMATYNOLSYOKAKYHLCGAGWLETR 111  
XX 32 eraagvyhrearsgkyklytaeakvcefgghlatykglearkigfhvcaagwmakgr 91  
XX  
XX 112 VAYPTAFASQNCGSGVGVIVDYGPENKSEMMDVFCYRMDVNC 155  
XX 92 vgyplvpgncpgcgtgldygrlnrserwdaycynphakec 135  
XX  
XX  
XX RESULT 3  
XX W84087  
XX ID W84087 standard; Protein; 277 AA.  
XX AC W84087;  
XX  
XX 10-FEB-1999 (first entry)  
XX  
XX Tumour necrosis factor stimulated gene 6 (TSG-6) protein.  
XX  
XX Tumour necrosis factor stimulated gene 6; TSG-6; recombinant; human;  
XX diagnosis; cytokine; immune disorder; autoimmune disorder; infection;  
XX inflammatory disease; neurodegenerative disease; cancer; hepatitis.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX Peptide 1..19  
XX Protein 20..277 /note= "putative signal sequence"  
XX Modified-site 118..120 /note= "mature protein"  
XX Modified-site 258..260 /note= "potential Asn-glycosylation site"  
XX /note= "potential Asn-glycosylation site"  
XX  
XX US5846763-A.  
XX  
XX 08-DEC-1998.  
XX  
XX 13-MAY-1994; 94US-0242097.  
XX  
XX 13-MAY-1994; 94US-0242097.  
XX  
XX 14-JAN-1991; 91US-0642312.  
XX  
XX 01-MAR-1993; 93US-0024868.  
XX  
XX (UUNY ) UNIV NEW YORK STATE.  
XX  
XX Lee TH, Vilcek J, Wisniewski H;  
XX









SEQ ID 11, Database: A-GeneSeq-36  
AC NO. W84087

File Copy

File Copy

PR 14-JAN-1991; 91US-0642312.

XX (UNY ) UNIV NEW YORK STATE.

XX Lee TH, Vilcek J, Wisniewski HG;

XX WPI; 1992-284330/34.

XX DR N-PSDB; Q27190.

XX Tumour necrosis factor-induced (glyco)protein mol. TSG-6 and its  
PT antibody - for treatment and diagnosis of chronic inflammatory  
PT conditions e.g. rheumatoid arthritis, infections, sepsis and  
PT cancer

XX Claim 1; Fig 3; 122pp; English.

XX The sequence is that of tumour necrosis factor (TNF) induced  
CC glycoprotein TSG-6. TSG-6 is induced in connective tissue by TNF  
CC and interleukin-7, and interferes with cell adhesion to their cells or  
CC the extracellular matrix. It and antibodies specific for it are useful  
CC in diagnosis (including identifying susceptibility) and treatment of  
CC various disorders associated with cytokine (in)activation e.g. chronic  
CC inflammation (esp. where associated with proteoglycan breakdown, such  
CC as rheumatoid arthritis) cancer (esp. development of metastases) and  
CC infections (esp. those caused by Gram negative bacteria).

XX Sequence 277 AA;

Query Match 14.1%; Score 262; DB 13; Length 277;

Best Local Similarity 43.3%; Pred. No. 1.6e-19;

Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

QY 52 DTTGVFHLRSPGQYKLTFRKAREACANEATMATYNQLSYXOKAKYHLCAGWLETGR 111

Db 32 eraagvyhrearsgkyktyeaakavcefegghlatyqleaarkgfhvcaagwmakgr 91

QY 112 VAYPTAFASONGSGVGVIVDYGPRNKSEMDVFCYRMKDVCNC 155

Db 92 vgyipvkpgncgfgktgildyginhrserwdacycynphakec 135

RESULT 2

W13654

ID W13654 standard; protein; 277 AA.

XX W13654;

XX 07-NOV-1997 (first entry)

XX Tumour necrosis factor-stimulated gene-6 protein.

XX TSG-6; human; tumour necrosis factor-stimulated gene-6 protein; treat;  
XX inflammatory disease; cancer-related pathology; inhibit; metastasis;  
XX invasive tumour growth; rheumatoid arthritis; cachexia; atherosclerosis;  
XX systemic lupus erythematosus; rational drug design.

XX Homo sapiens.

XX WO9704075-A1.

XX 06-FEB-1997.

XX 19-JUL-1996; 96WO-US11995.

XX 20-JUL-1995; 95US-0001311.

XX (UNY ) UNIV NEW YORK STATE.

XX Cronstein BN, Vilcek J, Wisniewski H;

XX WPI; 1997#132619/12.

PT New TSG-6 compositions to treat inflammatory diseases and cancer -  
PT can be complexed with inter-alpha-inhibitor, to treat auto-immune  
PT disease, cancer-related pathology etc. can also be administered as  
PT TSG-6 DNA

XX Claim 2; Page 58-59; 76pp; English.

XX This protein is the human tumour necrosis factor-stimulated gene-6  
CC protein (TSG-6). It can be used in a claimed pharmaceutical composition  
CC for treating inflammatory diseases and disorders, or cancer-related  
CC pathologies. The composition or DNA encoding TSG-6 may be used to inhibit  
CC tumour metastases, invasive tumour growth or cancer-related pathologies.  
CC They may also be used in the treatment of inflammatory conditions  
CC including rheumatoid arthritis, systemic lupus erythematosus, and  
CC infections such as cachexia, atherosclerosis etc. TSG-6 can also be used  
CC to design drugs to treat these disorders.

XX Sequence 277 AA;

Query Match 14.1%; Score 262; DB 18; Length 277;

Best Local Similarity 43.3%; Pred. No. 1.6e-19;

Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

QY 52 DTTGVFHLRSPGQYKLTFRKAREACANEATMATYNQLSYXOKAKYHLCAGWLETGR 111

Db 32 eraagvyhrearsgkyktyeaakavcefegghlatyqleaarkgfhvcaagwmakgr 91

QY 112 VAYPTAFASONGSGVGVIVDYGPRNKSEMDVFCYRMKDVCNC 155

Db 92 vgyipvkpgncgfgktgildyginhrserwdacycynphakec 135

RESULT 3

W84087

ID W84087 standard; Protein; 277 AA.

XX W84087;

XX 10-FEB-1999 (first entry)

XX Tumour necrosis factor stimulated gene 6 (TSG-6) protein.

XX Tumour necrosis factor stimulated gene 6; TSG-6; recombinant; human;  
XX diagnosis; cytokine; immune disorder; autoimmune disorder; infection;  
XX inflammatory disease; neurodegenerative disease; cancer; hepatitis.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..19

XX Protein /note= "putative signal sequence"

XX Modified-site 20..277

XX Modified-site /note= "mature protein"

XX Modified-site 118..120

XX Modified-site /note= "potential Asn-glycosylation site"

XX Modified-site 258..260

XX Modified-site /note= "potential Asn-glycosylation site"

XX US5846763-A.

XX 08-DEC-1998.

XX 13-MAY-1994; 94US-0242097.

XX 13-MAY-1994; 94US-0242097.

XX 14-JAN-1991; 91US-0642312.

XX 01-MAR-1993; 93US-0024868.

XX (UNY ) UNIV NEW YORK STATE.

XX Lee TH, Vilcek J, Wisniewski H;

XX WPI; 1997#132619/12.







OM protein - protein search, using sw model

Run on: April 4, 2001, 13:04:01 ; Search time 27.49 seconds  
(without alignments)  
1505.071 Million cell updates/sec

Title: US-09-466-778-11  
Perfect score: 1857  
Sequence: 1 MTGPGKHCKECSHYVGDL.....ALAAYSYFRINKTIGFXHF 353

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL15.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1757	94.6	897	4 Q9NRY3	Q9NRY3 homo sapien
2	1468	79.1	1069	4 Q9UF98	Q9UF98 homo sapien
3	662	35.6	2212	4 Q93072	Q93072 homo sapien
4	645.5	34.8	2570	4 Q9NY15	Q9NY15 homo sapien
5	264	14.2	275	11 Q08859	Q08859 mus musculus
6	177	9.5	2109	13 P79787	P79787 gallus gall
7	169.5	9.1	355	11 Q921X7	Q921X7 mus musculus
8	162	8.7	1321	4 Q14594	Q14594 homo sapien
9	161.5	8.7	655	11 Q08564	Q08564 rattus norv
10	161.5	8.7	1290	13 Q9W6E1	Q9W6E1 gallus gall
11	160.5	8.6	2394	6 Q77610	Q77610 bos taurus
12	153	8.2	192	6 Q02817	Q02817 oryctolagus
13	153	8.2	656	6 Q77612	Q77612 bos taurus
14	151.5	8.2	1643	6 Q77611	Q77611 bos taurus
15	151.5	8.2	3381	6 Q77609	Q77609 bos taurus
16	126.5	6.8	103	6 Q9TBB3	Q9TBB3 sus scrofa
17	120.5	6.5	103	6 Q46380	Q46380 oryctolagus
18	110	5.9	396	13 Q9W6S4	Q9W6S4 gallus gall
19	109	5.9	302	4 Q9UNF4	Q9UNF4 homo sapien

20	109	5.9	322	4 Q9Y5V7	Q9Y5V7 homo sapien
21	105	5.7	1584	5 Q93791	Q93791 caenorhabdi
22	104	5.6	3507	5 Q23587	Q23587 caenorhabdi
23	103.5	5.6	717	13 P87357	P87357 brachydanio
24	102.5	5.5	778	13 Q9IBG4	Q9IBG4 xenopus lae
25	100	5.4	816	11 Q70474	Q70474 rattus norv
26	99.5	5.4	682	4 Q9NTW2	Q9NTW2 homo sapien
27	98.5	5.4	728	13 Q90656	Q90656 gallus gall
28	99.5	5.4	802	13 Q57462	Q57462 brachydanio
29	99.5	5.4	3623	4 Q60494	Q60494 homo sapien
30	99	5.3	780	11 Q9QYX8	Q9QYX8 mus musculu
31	99	5.3	1440	5 Q20204	Q20204 caenorhabdi
32	99	5.3	2906	11 Q9WUJ9	Q9WUJ9 rattus norv
33	98	5.3	294	4 Q92493	Q92493 homo sapien
34	98	5.3	2531	5 Q16004	Q16004 lytechinus
35	97.5	5.3	780	11 Q08779	Q08779 rattus norv
36	97.5	5.3	1480	5 Q9V7F8	Q9V7F8 drosophila
37	97.5	5.3	1504	5 Q9XV4	Q9XV4 drosophila
38	97.5	5.3	1504	5 Q9V7F9	Q9V7F9 drosophila
39	96	5.2	364	11 Q70509	Q70509 rattus norv
40	96	5.2	742	4 Q9UJ36	Q9UJ36 homo sapien
41	95.5	5.1	364	6 Q97569	Q97569 ceratother
42	95.5	5.1	721	13 Q91902	Q91902 xenopus
43	95	5.1	261	10 Q24530	Q24530 vitis vinif
44	94.5	5.1	1328	13 P79754	P79754 fugu rubrip
45	94	5.1	637	10 Q9SF50	Q9SF50 arabidopsis

ALIGNMENTS

RESULT 1

Q9NRY3 ID Q9NRY3 PRELIMINARY; PRT; 897 AA.  
AC Q9NRY3;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE CD44-LIKE PRECURSOR FELL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tao Q., Zhang W., Cao X.;  
RT "Molecular cloning and characterization of human FELL sharing homology with CD44."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF160476; AAF62398.1; -  
SQ SEQUENCE 897 AA; 97585 MW; EB920AF36101E388 CRC64;

Query Match 94.6%; Score 1757; DB 4; Length 897;  
Best Local Similarity 93.8%; Pred. No. 8.6e-162;  
Matches 331; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

QY	1	MTGPGKHCKESHVYVGLNCEPEQLPDRCLQDNGOCHADAKCYDLHFQDTTGVFHL	60
DB	489	MTGPGKHCKESHVYVGLNCEPEQLPDRCLQDNGOCHADAKCYDLHFQDTTGVFHL	548
QY	61	RSPLQYKLTDFDKAREACANEAATMATYNQLSYQKAKYHLCAGWLETGRVAYPTAFAS	120
DB	549	RSPLQYKLTDFDKAREACANEAATMATYNQLSYQKAKYHLCAGWLETGRVAYPTAFAS	608
QY	121	QNCGSVVGVVDYGRPNKSEMDVFCYRMKDVMCTKVGKVGDFSGYSGNLQVLMSEFP	180
DB	609	QNCGSVVGVVDYGRPNKSEMDVFCYRMKDVMCTKVGKVGDFSGYSGNLQVLMSEFP	668
QY	181	SLTNELTEVLAYSNSARGRAFLHLDLSIRGTFLFPQNSGLGENETLSGRDIEHHLAN	240
DB	669	SLTNELTEVLAYSNSARGRAFLHLDLSIRGTFLFPQNSGLGENETLSGRDIEHHLAN	728



```
QY 241 VSMFYNDLVNGTTTQTRGLSKLLITRDQPLHPTETRCVGDGHDLEWDCASNGITHVI 300
DB 729 VSMFYNDLVNGTTTQTRGLSKLLITRDQPLHPTETRCVGDGHDLEWDCASNGITHVI 788
QY 301 SRXLKAPPAPVTLTHTXGLGXFXXIIIVTGAVALAAYSFRINRRTIGFQHF 353
DB 789 SRPLKAPPAPVTLTHTXGLGXFXXIIIVTGAVALAAYSFRINRRTIGFQHF 841
RESULT 2
ID Q9UF98 PRELIMINARY; PRT; 1069 AA.
AC Q9UF98
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHETICAL 115.7 KDA PROTEIN (FRAGMENT).
GN DKFZP434E0321.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID-9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133021; CAB61358.1;
DR HSP; P98066; ITSG.
DR INTERPRO; IPR000538;
DR INTERPRO; IPR000561;
DR INTERPRO; IPR000782;
DR INTERPRO; IPR002049;
DR PFAM; PF00008; EGF_8.
DR PFAM; PF00193; Xlink; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01265; LINKMODULE.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 5.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1069 AA; 115666 MW; 04B0960932164D63 CRC64;

Query Match 79.18; Score 1468; DB 4; Length 1069;
Best Local Similarity 80.5%; Pred. No. 1.2e-133;
Matches 284; Conservative 4; Mismatches 29; Indels 36; Gaps 2;

QY 1 MTGPKHKCKESHVYVDGLNCEPEQLPIDRCLODNGQCHADAKCVDLHFQDPTTVGVFHL 60
DB 697 MTGPKHKCKESHVYVDGLNCEPEQLPIDRCLODNGQCHADAKCVDLHFQDPTTVGVFHL 756
QY 61 RSLPGQYKLTDPKAREACANEAATMATYNQLSYQAKYHLCSSAGWLETGRVAYPTAFAS 120
DB 757 RSLPGQYKLTDPKAREACANEAATMATYNQLSYQAKYHLCSSAGWLETGRVAYPTAFAS 816
QY 121 QNCGSGVGVYDYGPRPNKSEMDVFCYRMKDVNCTKXKVGVDGFSYSGNLLQVLMSP 180
DB 817 QNCGSGVGVYDYGPRPNKSEMDVFCYRMKDVNCTKXKVGVDGFSYSGNLLQVLMSP 860
QY 181 SLTNFLTEVLAYSNSARGRAFLHLEHLDLSIRGTLFQPNQSGENETLSGRDIEHHLA 240
DB 861 SLTNFLTEVLAYSNSARGRAFLHLEHLDLSIRGTLFQPNQSGENETLSGRDIEHHLA 900
QY 241 VSMFYNDLVNGTTTQTRGLSKLLITRDQPLHPTETRCVGDGHDLEWDCASNGITHVI 300
DB 901 VSMFYNDLVNGTTTQTRGLSKLLITRDQPLHPTETRCVGDGHDLEWDCASNGITHVI 960
QY 301 SRXLKAPPAPVTLTHTXGLGXFXXIIIVTGAVALAAYSFRINRRTIGFQHF 353
DB 961 SRPLKAPPAPVTLTHTXGLGXFXXIIIVTGAVALAAYSFRINRRTIGFQHF 1013
```

```
RESULT 3
ID Q93072 PRELIMINARY; PRT; 2212 AA.
AC Q93072
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE MREDOBLAST KIAA0246 PROTEIN (FRAGMENT).
GN KIAA0246.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID-9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BONE MARROW;
RX MEDLINE-97191844; PubMed-9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.
RL DNA Res. 3:321-329 (1996).
DR EMBL; D87433; BAA1337.1;
DR HSP; P98066; ITSG.
DR INTERPRO; IPR000538;
DR INTERPRO; IPR000561;
DR INTERPRO; IPR000782;
DR INTERPRO; IPR001128;
DR INTERPRO; IPR002049;
DR PFAM; PF00008; EGF_13.
DR PFAM; PF00193; Xlink; 1.
DR PRINTS; PR01265; LINKMODULE.
DR PROSITE; PS00022; EGF_1; UNKNOWN_5.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 13.
DR PROSITE; PS01241; LINK; UNKNOWN_1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 2.
KW Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 2212 AA; 237451 MW; 4A95480504129134 CRC64;

Query Match 35.6%; Score 662; DB 4; Length 2212;
Best Local Similarity 40.2%; Pred. No. 4.9e-55;
Matches 143; Conservative 47; Mismatches 158; Indels 8; Gaps 5;

QY 2 TGPGKHKCKESHVYVDGLNCEPEQLPIDRCLODNGQCHADAKCVDLHFQDPTTVGVFHL 60
DB 1796 TGLNTRCRCHAGYVDGLQCLESEPPVDRCLQGPCHSDANCTDLHFQKRGVGH 1855
QY 61 RSLPGQYKLTDPKAREACANEAATMATYNQLSYQAKYHLCSSAGWLETGRVAYPTAFAS 120
DB 1856 QATSGPYGLNFEAEAEACAGAVLASEPQLSAAQQLGFHLCMLGALANGSTAHVPVFPV 1915
QY 121 QNCGSGVGVYDYGPRPNKSEMDVFCYRMKDVNCTKXKVGVDGFSYSGNLLQVLMSP 179
DB 1916 ADCNGRGVIGVLSGAKNLSERDAICFRQVDVACRNGFVGDIETCKKLDVLAAT 1975
QY 180 SLTNFLTEVLAYSNSARGRAFLHLEHLDLSIRGTLFQPNQSGENETLSGRDIEHHLA 239
DB 1976 ANFSTFVGLLSVANATQRGIDFLDDELTYKTFLFPVNEGFVDNMTLSGPNLEHAS 2035
QY 240 VSMFYNDLVNGTTTQTRGLSKLLITDR---QDPLHPTETRCVGDGHDLEWDCASNGI 296
DB 2036 NATLLSAN-ASQCKLLPAHSGLSLIISDAGDPDSSWAPAGPTVVVSRIIVDMIFNGI 2094
QY 297 TVHSRLKAPPAPVTLTHTXGLGXFXXIIIVTGAV--ALAAYSYFRINRRTIGF 350
DB 2095 IHALASPLAPQPAVLAPAEAPPVAGVLAAGALLGVAGALYLRARGKPMGF 2150
```

